

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 12, 2003, 02:59:14 ; Search time 80 Seconds
(without alignments)
5128.796 Million cell updates/sec

Title: US-09-617-099b-1

Perfect score: 8285

Sequence: 1 MSAPLGPRGRPAPTAAQP.....TTRASQSSLESSTGSPSYSES 1590

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

- 1: sp archaea:*
- 2: sp bacteria:*
- 3: sp fungi:*
- 4: sp human:*
- 5: sp invertebrate:*
- 6: sp mammal:*
- 7: sp mhc:*
- 8: sp organelle:*
- 9: sp phage:*
- 10: sp plant:*
- 11: sp rodent:*
- 12: sp virus:*
- 13: sp vertebrate:*
- 14: sp unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7930	95.7	1530	11 Q9EQZ7	Q9eqz7 mus musculu
2	7459.5	90.0	1525	11 Q9JIR5	Q9jir5 rattus norv
3	7438.5	89.8	1539	11 Q9JIR6	Q9jir6 rattus norv
4	7420.5	89.6	1555	11 Q9JIR1	Q9jir1 rattus norv
5	6856.5	82.8	1399	11 Q9JIS0	Q9jis0 rattus norv
6	6543	79.0	1330	11 Q9JIR7	Q9jir7 rattus norv
7	6522	78.7	1352	11 Q9JIR6	Q9jir6 rattus norv
8	6504	78.5	1368	11 Q9JIR9	Q9jir9 rattus norv
9	6261.5	75.6	1363	11 Q8CIR2	Q8cix2 rattus norv
10	6212	75.0	1292	11 Q9JIR8	Q9jir8 rattus norv
11	5638	68.1	1188	4 Q9JQ26	Q9jq26 homo sapien
12	4513.5	54.5	1615	11 Q9JIR4	Q9jir4 rattus norv
13	4489.5	54.2	1553	11 Q35168	Q35168 rattus norv
14	4402	53.1	1470	4 Q8TDY9	Q8tdy9 homo sapien
15	2936.5	35.4	1028	4 Q8TD25	Q8td25 homo sapien
16	2936.5	35.4	1053	4 O15048	O15048 homo sapien

Search before
Allowance

17	2424.5	29.3	740	4 Q9HBA5	Q9hba5 homo sapien
18	2410.5	29.1	766	4 Q9HBA6	Q9hba6 homo sapien
19	2340.5	28.2	700	4 Q9HBA4	Q9hba4 homo sapien
20	1912	23.1	369	4 O43413	O43413 homo sapien
21	1834.5	22.1	596	4 Q9HBA3	Q9hba3 homo sapien
22	1821	22.0	567	4 Q9HBA2	Q9hba2 homo sapien
23	1763.5	21.3	516	4 Q9HBA1	Q9hba1 homo sapien
24	1622	19.6	352	11 Q8C433	Q8c433 mus musculu
25	1507.5	18.2	471	4 Q9NUB3	Q9nub3 mus musculu
26	1452.5	17.5	1563	5 Q22366	Q22366 caenorhabdi
27	1224.5	14.8	285	11 Q8CCK2	Q8ck2 mus musculu
28	1221.5	14.7	285	11 Q9JIR2	Q9jir2 rattus norv
29	1041	12.6	1371	5 Q8MPX5	Q8mpx5 caenorhabdi
30	1025.5	12.4	308	4 Q9UUD0	Q9ujd0 homo sapien
31	1022	12.3	307	11 Q9JIR3	Q9jir3 rattus norv
32	1021.5	12.3	308	4 Q92511	Q92511 homo sapien
33	987.5	11.9	2469	5 Q9VEE5	Q9vee5 drosophila
34	841.5	10.2	268	4 Q9H426	Q9h426 homo sapien
35	834	10.1	388	11 Q99NE5	Q99ne5 mus musculu
36	825	10.0	269	11 Q8CIX1	Q8cix1 rattus norv
37	623	7.5	136	4 Q8IWW9	Q8iww9 homo sapien
38	554	6.7	5165	11 Q8CF92	Q8cf92 mus musculu
39	459.5	5.5	4969	11 Q8CF91	Q8cf91 mus musculu
40	454	5.5	1592	11 Q8CH28	Q8che28 mus musculu
41	443	5.3	1231	5 Q9BIE6	Q9bi66 caenorhabdi
42	437	5.3	113	4 Q8IWW1	Q8iww1 homo sapien
43	437	5.3	126	4 Q8IWW0	Q8iww0 homo sapien
44	371	4.5	2296	4 Q9UHA8	Q9uha8 homo sapien
45	369	4.5	2752	4 Q9UQ35	Q9uq35 homo sapien

ALIGNMENTS

RESULT 1	Q9EQZ7	PRELIMINARY;	PRT; 1530 AA.
ID	Q9EQZ7		
AC	Q9EQZ7		
DT	01-MAR-2001 (TRENBLrel. 16, Created)		
DT	01-MAR-2001 (TRENBLrel. 16, Last sequence update)		
DT	01-MAR-2003 (TRENBLrel. 23, Last annotation update)		
DE	Rim2		
GN	RIMS2 OR RAB3IP2 OR RIM2.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEBLINE=20512528; PubMed=11056535;		
RA	Ozaki N., Shibasaki T., Kashima Y., Miki T., Takahashi K., Ueno H.,		
RA	Sunaga Y., Yano H., Matsuura Y., Iwanaga T., Takai Y., Seino S.;		
RT	"CAMP-GEPII is a direct target of CAMP in regulated exocytosis."		
RL	Nat. Cell Biol. 2:805-811(2000)		
CC	-1- SIMILARITY: CONTAINS 2 C2 DOMAINS.		
CC	-1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.		
DR	EMBL; AB021131; BAB18975.1; --		
DR	HSSP; P21707; 1BYN.		
DR	MGP; MGI:2152972; Rims2.		
DR	InterPro; IPR000008; C2.		
DR	InterPro; IPR000345; CytC_heme_bind.		
DR	InterPro; IPR001478; PDZ.		
DR	InterPro; IPR00306; Znf_FYVE.		
DR	Pfam; PF00168; C2; 2.		
DR	Pfam; PF00595; PDZ; 1.		
DR	SMART; SM00239; C2; 2.		
DR	SMART; SM00228; PDZ; 1.		
DR	PROSITE; PS50004; C2 DOMAIN 2; 2.		
DR	PROSITE; PS00190; CYTOCHROME_C; 1.		
DR	PROSITE; PS0106; PDZ; 1.		
DR	PROSITE; PS50178; ZF FYVE; 1.		
DR	SEQUENCE 1530 AA; 172863 MW; 58CF11BF7152357D CRC64;		

Query Match		95.7%; Score 7930; DB 11; Length 1530;
Best Local Similarity		96.2%; Pred. No. 0;
Matches 1529; Conservative		0; Mismatches 1; Indels 60; Gaps 1;
QY	1	MSAPLGRGRGAPAPPAASQPPPOPEMDLHSLTEERKIIILAVMDRQKEEKEEQSVLKI 60
DB	1	MSAPLGRGRGAPAPPAASQPPPOPEMDLHSLTEERKIIILAVMDRQKEEKEEQSVLKI 60
QY	61	KEBHKAQPTQWFFPSGITELVNNVLOPQOQPNKEKEPQTKLHQOFENYKEQVKKWGESQ 120
DB	61	KEBHKAQPTQWFFPSGITELVNNVLOPQOQPNKEKEPQTKLHQOFENYKEQVKKWGESQ 120
QY	121	QOEOQKGDAPTCGICHTKTKADCGCHNCSCYQTFPCARCGRVSLSRNKVMVNCICRQ 180
DB	121	QOEOQKGDAPTCGICHTKTKADCGCHNCSCYQTFPCARCGRVSLSRNKVMVNCICRQ 180
QY	181	QEILTKSGANFYNSGNTLQOPQKVPGRGLNREAPQEKAKLHBOPOFOGAPCDLSVPA 240
DB	181	QEILTKSGANFYNSGNTLQOPQKVPGRGLNREAPQEKAKLHBOPOFOGAPCDLSVPA 240
QY	241	VEKRAHGLTRQDTIKNGSGVGHQIASDMPSPDRKSPSVSRDQNRREYQSEEREDYSQYV 300
DB	241	VEKRAHGLTRQDTIKNGSGVGHQIASDMPSPDRKSPSVSRDQNRREYQSEEREDYSQYV 300
QY	301	PSDGTWPRSPSDYADRSPQEPQYEPHLYRDSNRGRHSKEYIYVDDDEVSRDEY 360
DB	301	PSDGTWPRSPSDYADRSPQEPQYEPHLYRDSNRGRHSKEYIYVDDDEVSRDEY 360
QY	361	ERQRREVEYQARYSRDPNLARYPVKQPYEBOIRIAEVSRAHRRHSDVSLANAELED 420
DB	361	ERQRREVEYQARYSRDPNLARYPVKQPYEBOIRIAEVSRAHRRHSDVSLANAELED 420
QY	421	SRISLLMRDRPSQRVSERRAAENORSYMERETREAQOQSSYPQRTSNHSPPTPRSP 480
DB	421	SRISLLMRDRPSQRVSERRAAENORSYMERETREAQOQSSYPQRTSNHSPPTPRSP 480
QY	481	IPLDPRDMRADSLRQKHLDPSSAVRKYTRKQETMLNDSLSQSSVRPPPPRPHK 540
DB	481	IPLDPRDMRADSLRQKHLDPSSAVRKYTRKQETMLNDSLSQSSVRPPPPRPHK 540
QY	541	SKKGGKMRQVSLSSSEELASTPEYSCDDVLESSESVSEKGDQSKKRTSQGVLSDS 600
DB	541	SKKGGKMRQVSLSSSEELASTPEYSCDDVLESSESVSEKGDQSKKRTSQGVLSDS 600
QY	601	NTRSERQKRMYYGHSLEEDLSEWSPQIKDGVDTCSSTTLNEESHSDKHPVTWQPSK 660
DB	601	NTRSERQKRMYYGHSLEEDLSEWSPQIKDGVDTCSSTTLNEESHSDKHPVTWQPSK 660
QY	661	DGDLITGRILLNKLKXGSPRDSGAMGLKVVGKQMTESGRICAPITKVKKGLADTVG 720
DB	661	DGDLITGRILLNKLKXGSPRDSGAMGLKVVGKQMTESGRICAPITKVKKGLADTVG 720
QY	721	HLRPGDEVLEWNGRLQAGATFRVYNIILSKPEPOVELVSRPIGDIIPRIPDSHAQLE 780
DB	721	HLRPGDEVLEWNGRLQAGATFRVYNIILSKPEPOVELVSRPIGDIIPRIPDSHAQLE 780
QY	781	SSSSSFESQMDRPSISVTPSPSGMLRDVPQSLQSLIKLFWKVGHLVITILGAD 840
DB	781	SSSSSFESQMDRPSISVTPSPSGMLRDVPQSLQSLIKLFWKVGHLVITILGAD 840
QY	841	LPSREGRPNPYKLYFLPDRSDKDKRTKTKTKLTPNNQTFIYSPVHRERFERML 900
DB	841	LPSREGRPNPYKLYFLPDRSDKDKRTKTKTKLTPNNQTFIYSPVHRERFERML 900
QY	901	EITLMDQARVRESEFLEGLILELTALDDPHWVGLQTHDVSLPLPRSPYLPQ 960
DB	901	EITLMDQARVRESEFLEGLILELTALDDPHWVGLQTHDVSLPLPRSPYLPQ 960
QY	961	LHGESPTRELQSKRISDSSEVSDYCDGCVGVVSDYRHNDRQLQSTLSVPQVWSSNHC 1020
DB	961	LHGESPTRELQSKRISDSSEVSDYCDGCVGVVSDYRHNDRQLQSTLSVPQVWSSNHC 1020
QY	1021	SFSGSPHRVDVIGTRTSWSPSPAPPQPNVEQGHGRTATGHYNTISRMDRHRVMDHYSS 1080
DB	1021	SFSGSPHRVDVIGTRTSWSPSPAPPQPNVEQGHGRTATGHYNTISRMDRHRVMDHYSS 1080
QY	1081	DEDDCEADRDQPHRSRSTEQRPILLRTTTRSSSRSDPTNLRSMPSLMTORSAPPSP 1140
DB	1081	DEDDCEADRDQPHRSRSTEQRPILLRTTTRSSSRSDPTNLRSMPSLMTORSAPPSP 1140
QY	1141	ALRSRSHPTSGSVQTSPTSGTERRGRQLPOLPKGKTLERSAMDIERNRQMKLNYKQV 1200
DB	1084	---RSHPTSGSVQTSPTSGTERRGRQLPOLPKGKTLERSAMDIERNRQMKLNYKQV 1140
QY	1201	AGSDPRLRDYHSHKYSKSGWDPHRGADTVSTKSSDSVDSVSAVSRSTSSASRFSSTYSMV 1260
DB	1141	AGSDPRLRDYHSHKYSKSGWDPHRGADTVSTKSSDSVDSVSAVSRSTSSASRFSSTYSMV 1200
QY	1261	QSEEPGRNKEISVFTSKMQRONGVSGKULTKSTSTISGDMCSLEKNDQSDTAVALGT 1320
DB	1201	QSEEPGRNKEISVFTSKMQRONGVSGKULTKSTSTISGDMCSLEKNDQSDTAVALGT 1260
QY	1321	SGKGRSSIGAKWVAIVGLSRKSRASOLSQTEGGGKGLRSTTVORSTETGLAVENWMT 1380
DB	1261	SGKGRSSIGAKWVAIVGLSRKSRASOLSQTEGGGKGLRSTTVORSTETGLAVENWMT 1320
QY	1381	QASRESTGSMNSYSSSEGNLIPPGVRLASDSQPSDFLXLGPAQLVGRQTLATPAKGI 1440
DB	1321	QASRESTGSMNSYSSSEGNLIPPGVRLASDSQPSDFLXLGPAQLVGRQTLATPAKGI 1380
QY	1441	QYGMMDKGLQELVEIRARGLVVKQSKTLPAPYVVKVLLDNGVCIKKTUKVARTLEP 1500
DB	1381	QYGMMDKGLQELVEIRARGLVVKQSKTLPAPYVVKVLLDNGVCIKKTUKVARTLEP 1440
QY	1501	LYQQLLSFEESQGRVLIQIIVMGDVRMDHKSFMGVAQIILDELHLSNMVIGWFKLPPPS 1560
DB	1441	LYQQLLSFEESQGRVLIQIIVMGDVRMDHKSFMGVAQIILDELHLSNMVIGWFKLPPPS 1500
QY	1561	SLVDPSTAPLTPRASQSSLESSTGSPYSRS 1590
DB	1501	SLVDPSTAPLTPRASQSSLESSTGSPYSRS 1530
RESULT 2		
Q9JIR5		PRELIMINARY; PRT; 1525 AA.
ID	Q9JIR5	
AC	Q9JIR5	01-OCT-2000 (TrEMBLrel. 15, Created)
DT	01-OCT-2000	(TrEMBLrel. 15, Last sequence update)
DT	01-OCT-2000	(TrEMBLrel. 15, Last sequence update)
DT	01-WAR-2003	(TrEMBLrel. 23, Last annotation update)
DE	RIM2-5B.	
GN	RIM2-5B.	
OS	Rattus norvegicus (Rat).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	
OX	NCBI_Taxid=10116;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=20347919; PubMed=10748113;	
RA	Wang Y., Sugita S., Sudhof T.C.;	
RT	"The RIM/NIM Family of Neuronal C2 Domain Proteins. Interactions with	
RT	Rab3 and a new class of Src homology 3 domain proteins."	
RL	J. Biol. Chem. 275:20033-20044 (2000).	
CC	- - SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.	
CC	- - SIMILARITY: CONTAINS 2 C2 DOMAINS.	
DR	EMBL; AF199331; AAF81653.1; -	
DR	HSSP; P21707; 1BYN.	
DR	InterPro; IPR000008; C2.	
DR	InterPro; IPR000345; CytC_heme_bind.	
DR	InterPro; IPR001478; PDZ.	
DR	InterPro; IPR003315; RPH3A_effector.	
DR	Pfam; PF00168; C2; 2.	
DR	Pfam; PF00595; PDZ; 1.	
DR	Pfam; PF02318; RPH3A_effector; 1.	
DR	SMART; SM00239; C2; 2.	

DR SMART: SM00228; PDZ; 1.
 DR PROSITE; PS50004; C2 DOMAIN 2; 2.
 DR PROSITE; PS00190; C2 CYTOCHROME C; 1.
 DR PROSITE; PS0106; PDZ; 1.
 DR PROSITE; PS0178; 2F FYVE; 1.
 SQ SEQUENCE 1525 AA; 172435 MW; D4930D6D2F03CF21 CRC64;

Query Match 90.0%; Score 7459.5; DB 11; Length 1525;
 Best Local Similarity 90.1%; Pred. No. 0;
 Matches 1452; Conservative 21; Mismatches 30; Indels 109; Gaps 4;

QY 1 MSALPGRGRAPAPPAASOPPPQPPMPDLSHLTREERKIILAVMDROKKEEKEOSVLKI 60
 DB 1 MSALPGRGRAPAPPAASOPPPQPPMPDLSHLTREERKIILAVMDROKKEEKEOSVLK- 59
 QY 61 KEERKAQPTWFFPGITELVNNVLPQOKQPNKEKPTKLLHQPFMYKQVKMGESQ 120
 DB 60 -----KLHQPFMYKQVKMGESQ 80
 QY 121 QOQKQGDAPCGI CHTKFADGCHNCYCQTFCARCGGRVSLRNKVMWVNLCKQ 180
 DB 81 QOQKQGDAPCGI CHTKFADGCHNCYCQTFCARCGGRVSLRNKVMWVNLCKQ 140
 QY 181 QEILTKSGAFYNGSNTLQPDQKVPRLNREAPQEKAKLHQPQFOGAPGDLVPA 240
 DB 141 QEILTKSGAFYNGSNTLQPDQKALRLGKSEAPQEKAKLHQPQFOGAPGDLVPA 200
 QY 241 VEKRAHGLTRQDTIKNGSGVHQIADMPDRKSPSVSDQNRVYQSEEREDYQYV 300
 DB 201 VERGRAHGLTRQDTIKNGSGVHQIADMPDRKSPSVSDQNRVYQSEEREDYQYV 260
 QY 301 PSDGTPRSDYADRSQREPPQYEPFGHLYNDSNRGRHRSKYIVDDDEVSRDEY 360
 DB 261 PSDGTPRSDYADRSQREPPQYEPFGHLYNDSNRGRHRSKYIVDDDEVSRDEY 320
 QY 361 ERQRREYQARTSDNLARYPVKQPYEOMRIHAEVSPARHERHSDVSLANAELED 420
 DB 321 ERQRREYQARTSDNLARYPVKQPYEOMRIHAEVSPARHERHSDVSLANAELED 380
 QY 421 SRISLLRMDPRQRYSERRAAMENORSWERTREAOQSSYPQRTSNHSPPTPRSP 480
 DB 381 SRISLLRMDPRQRYSERRAAMENORSWERTREAOQSSYPQRTSNHSPPTPRSP 440
 QY 481 IPLDRPMRADSLRKQHLDPSSAVRTKREKMETMLRNDLSQDSQSESVRPPRPHK 540
 DB 441 IPLDRPMRADSLRKQHLDPSSAVRTKREKMETMLRNDLSQDSQSESVRPPRPHK 500
 QY 541 SKKGKMRQVSLSSSEELASTPEYTSRDDVLESESVSEKGSQKRTTSQGVLSDS 600
 DB 501 SKKGKMRQVSLSSSEELASTPEYTSRDDVLESESVSEKGSQKRTTSQGVLSDS 560
 QY 601 NTRSERQKRMYYGGHSLSELEDSSEPOIKDQSGVDTCSTTLNEEHSKDKHPVWQPSK 660
 DB 546 -----YSWLEHSAWH -----SSEASPMSLHPVWQPSK 573
 QY 661 DGRLLIGRIILNKLKDGSPVPSGAMGLKVGGKMTESGRLCFAFTKVKKSLADTVG 720
 DB 574 DGRLLIGRIILNKLKDGSPVPSGAMGLKVGGKMTESGRLCFAFTKVKKSLADTVG 633
 QY 721 HLRPGDEVLEWGRLLQGAFTFEVYNIILLESKPEQVELVVRPDIPIPISTHAQLE 780
 DB 634 HLRPGDEVLEWGRLLQGAFTFEVYNIILLESKPEQVELVVRPDIPIPISTHAQLE 693
 QY 781 SSSSSFSQMDRPSISVTPGMPGLRDVDPQLSGQLSLKLPDKVGHOLIIVTILGAKD 840
 DB 694 SSSSSFSQMDRPSISVTPGMPGLRDVDPQLSGQLSLKLPDKVGHOLIIVTILGAKD 753
 QY 841 LPSRDEGRPNPVKLYFLPDRSDKNKRTTKVKKTLKPKWNTFIYSPVHREFEERML 900
 DB 754 LPSRDEGRPNPVKLYFLPDRSDKNKRTTKVKKTLKPKWNTFIYSPVHREFEERML 813
 QY 901 EITLWQARVRESESEFLGILLLELTALLDDPHYKLTQTHDVSSLPDPSPYLPRRQ 960
 DB -----

814 EITLWQARVRESESEFLGILLLELTALLDDPHYKLTQTHDVSSLPDPSPYLPRRQ 873
 QY 961 LRGSPTRRLQRSKRIISDSVSDYDCEDGVGVSDYEHKGRDLQSSITLVPYEQVMSNHC 1020
 DB 874 LRGSPTRRLQRSKRIISDSVSDYDCEDGVGVSDYEHKGRDLQSSITLVPYEQVMSNHC 933
 QY 1021 SPGSPHRVDVIGTRTSWSPSPAPPQPNVQCEHGRGTRATGHYNTISMDHRVMDHYSS 1080
 DB 934 SPGSPHRVDVIGTRTSWSPSPAPPQPNVQCEHGRGTRATGHYNTISMDHRVMDHYSS 993
 QY 1081 DRD-----RDCEAADRQPHRSRSTQRPLLETTTTSRSESR 1118
 DB 994 ERDSHFLTLPRSRHROTSEHHRDGRDCEAADRQPHRSRSTQRPLLETTTTSRSESR 1053
 QY 1119 PDTNLARSMPLMTGRSAPPSPALSRSHPTGVSQTSPTSSTGTCTGREGROLPOLPPKGTIL 1178
 DB 1054 ADTNLARSMPLMTGRSAPPSPALSRSHPTGVSQTSPTSSTGTCTGREGROLPOLPPKGTIL 1113
 QY 1179 ERGAMDIERNRQMKLNKYQVAGSDPRLPDYTHSKYRSCWDPERGADTVSTKSSDSVS 1238
 DB 1114 ERGAMDIERNRQMKLNKYQVAGSDPRLPDYTHSKYRSCWDPERGADTVSTKSSDSVS 1173
 QY 1239 DVSASVSTSSASPSSTSYMSVQSERPRGRKLSVFTSRKMNQMGVSGKLTSTISIG 1298
 DB 1174 DVSASVSTSSASPSSTSYMSVQSERPRGRKLSVFTSRKMNQMGVSGKLTSTISIG 1233
 QY 1299 DMCLEKNDQSQSDTAVAGLTSGKRRSSIGAKWALVCLSRKSRBASQLSQTEGQKK 1358
 DB 1234 DMCLEKNDQSQSDTAVAGLTSGKRRSSIGAKWALVCLSRKSRBASQLSQTEGQKK 1293
 QY 1359 LRSTVQRSTETGLAVENRNMTRQASRESTDGSMNSYSSEGNLIFPGVRLASDQPSDFL 1418
 DB 1294 LRSTVQRSTETGLAVENRNMTRQASRESTDGSMNSYSSEGNLIFPGVRLASDQPSDFL 1353
 QY 1419 DGLGPAQLVGRQTLATPAMCDIQVGMCKGQLEVEIIRARGLVVTPGSKTLPATYKVV 1478
 DB 1354 DGLGPAQLVGRQTLATPAMCDIQVGMCKGQLEVEIIRARGLVVTPGSKTLPATYKVV 1413
 QY 1479 LLDNGVCIANKKTKVARKTLEPLYQQLLSFEESPOGVLOIIVMGDYGMDHESFMVAQ 1538
 DB 1414 LLDNGVCIANKKTKVARKTLEPLYQQLLSFEESPOGVLOIIVMGDYGMDHESFMVAQ 1473
 QY 1539 ILDALELSNMVGNFKLPPPSLVDPSTAPLTRASQSSLESSTQPSYRS 1590
 DB 1474 ILDALELSNMVGNFKLPPPSLVDPSTAPLTRASQSSLESSTQPSYRS 1525

RESULT 3
 Q9JIR6
 ID Q9JIR6 PRELIMINARY; PRT; 1539 AA.
 AC Q9JIR6;
 DT 01-OCT-2000 (T-EMBLrel. 15, Created)
 DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)
 DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
 DE RIM2-5A.
 GN RIM2-5A.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=101116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20347919; PubMed=10748113;
 RA Wang Y., Sugita S., Sudhof T.C.;
 RT "The RIM/NIM Family of Neuronal C2 Domain Proteins. Interactions with
 RT Rab3 and a new class of Src homology 3 domain proteins.";
 RL J. Biol. Chem. 275:20033-20044(2000).
 CC -1- SIMILARITY: CONTAINS 2 C2 DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
 DR EMBL; AF199330; AAF81652.1; --
 DR HSP; P21707; 1BYN.
 DR InterPro; IPR000008; C2.
 DR InterPro; IPR000345; CytC_heme_bind.

DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE RIM2-2B.
 GN RIM2-2B.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20347919; PubMed=10748113;
 RA Wang Y., Sugita S., Sudhof T.C.;
 RT "The RIM/NIM Family of Neuronal C2 Domain Proteins. Interactions with
 RT Rab3 and a new class of Src homology 3 domain proteins.";
 RL J. Biol. Chem. 275:20033-20044(2000).
 CC 1- SIMILARITY: CONTAINS 2 C2 DOMAINS.
 CC 1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
 DR EMBL; AF193324; AAF81646.1; -;
 DR HSP; F21707; IBYN.
 DR InterPro; IPR000008; C2.
 DR InterPro; IPR000345; CytC_heme_bind.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR003315; RPH3A_effector.
 DR InterPro; IPR000306; Znf_FYVE.
 DR Pfam; PF00168; C2; 2.
 DR Pfam; PF00595; PDZ; 1.
 DR SMART; SM00238; PDZ; 1.
 DR SMART; SM00239; C2; 2.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS50004; C2_DOMAIN_2; 2.
 DR PROSITE; PS50019; CYTOCHROME_C; 1.
 DR PROSITE; PS50106; PDZ; 1.
 DR PROSITE; PS50178; ZF_FYVE; 1.
 DR PROSITE; PS50178; ZF_FYVE; 1.
 SQ SEQUENCE 1399 AA; 158945 MW; 3F1550E3D52D237B CRC64;

 Query Match 82.8%; Score 6856.5; DB 11; Length 1399;
 Best Local Similarity 83.0%; Pred. No. 0;
 Matches 1338; Conservative 17; Mismatches 22; Indels 235; Gaps 3;

 QY 1 MSAPLGRPRGAPPTAASPPPPPEMPDLISLHTEERKLIILAVMDRKKERKEQSVLKI 60
 DB 1 MSAPLGRPRGAPPTAASPPPPPEMPDLISLHTEERKLIILAVMDRKKERKEQSVLKI- 59
 QY 61 KEEHKAQPTQWPPFGITELVANNVLPQOQKPNKEPOTKLHQQFEMYKQVKKGERSQ 120
 DB 60 -----KLHQQFEMYKQVKKGERSQ 80
 QY 121 QOQEQKGDAPTCGI CHKTFADCGCHNCSCYQTKFCARCGGRVSLRSNKNVWVCMCRKQ 180
 DB 81 QOQEQKGDAPTCGI CHKTFADCGCHNCSCYQTKFCARCGGRVSLRSNKNVWVCMCRKQ 140
 QY 181 QILYTKSGAFWYNSGNTLQPDQKVPGLRNEEAPQEKKALHEQPOFGAPGDLSPVA 240
 DB 141 QILYTKSGAFWYNSGNTLQPDQKALGLRSEAPQEKKALHEQTQFGPPGSSVPA 200
 QY 241 VEKGRAGHLTROPTINGSGVHQIASDMPDRKSPSVSRQNRRYEGEREYSQYV 300
 DB 201 VERGRAGHLTRQDSIKNGSGMKHQIASDMPDRKSPSVSRQNRRYEGEREYSQYV 260
 QY 301 PSDGTMPSPDYARRSOREPOFEEPCGHINVRNRRGHRSKEYIYDDDEDVSRDEY 360
 DB 261 PSDTMTSPSPDYARRSOREPOFEEPCGHINVRNRRGHRSKEYIYDDDEDVSRDEY 320
 QY 361 ERQREERYQARYSDPNLARVPVQPOFEEQWRTHAEVSRARHRRHSDVSLANAELED 420
 DB 321 ERQREERYQARYSDPNLARVPVQPOFEEQWRTHAEVSRARHRRHSDVSLANAELED 380
 QY 421 SRISLLRMDRPSRQSVSERRAAMENQRSYSMERTREAGQSSYFQRTSNHSPPTPRSP 480
 DB 381 SRISLLRMDRPSRQSVSERRAAMENQRSYSMERTREAGQSSYFQRTSNHSPPTPRSP 440
 QY 481 IFLDRPDMRRADSLRKQHLDPSSAVRKYTKREKMETMLRNDLSLSDQSVSRPPPRPHK 540

DB 441 IFLDRPDMRRADSLRKQHLDPSSAVRKYTKREKMETMLRNDLSLSDQSVSRPPPRPHK 500
 QY 541 SKGGMEQVLSSESEELASTPYTSCDDVLESESVSEKGDQKGRKTSQGVLSDS 600
 DB 501 SKGGMEQVLSSESEELASTPYTSCDDVLESESVSEKGDQKGRKTSQGVLSDS 560
 QY 601 NTRSERQKRMYYGGHSLSEEDLEWSEPOIKDQGVDTCSSTTLNEHSHSDKHPVTPQSK 660
 DB 561 NTRSERQKRMYYGGHSLSEEDLEWSEPOIKDQGVDTCSSTTLNEHSHSDKHPVTPQSK 620
 QY 661 DQDRLLIGRILLNKRLLKQSVPRDSGAMGLKVGGKMTESGRCLAFITVYKGSADTVG 720
 DB 621 DQDRLLIGRILLNKRLLKQSVPRDSGAMGLKVGGKMTESGRCLAFITVYKGSADTVG 680
 QY 721 HLRPDQVLEWNGRLLOQATFEVYNIILESKPEQVLEVSRPITGDIPLIPDSHAQLE 780
 DB 681 HLRPDQVLEWNGRLLOQATFEVYNIILESKPEQVLEVSRPITGDIPLIPDSHAQLE 740
 QY 781 SSSSFESQKMDRPSISVTSFMSQMLRDVFPQFLSGQLSIKLWFDKVGHLIVITLGAKD 840
 DB 741 SSSSFESQKMDRPSISVTSFMSQMLRDVFPQFLSGQLSIKLWFDKVGHLIVITLGAKD 800
 QY 841 LPSREDGEPRPYVYKIYFLPDRSDKGRITVYKTLKPKNQVTFITFVRRERPRML 900
 DB 801 LPSREDGEPRPYVYKIYFLPDRSDKGRITVYKTLKPKNQVTFITFVRRERPRML 860
 QY 901 EITLWDQARVRESESEFLGELILETALLDDEPHYKLTQTHDVSSLPPLPPSPVLRQ 960
 DB 861 EITLWDQARVRESESEFLGELILETALLDDEPHYKLTQTHDVSSLPPLPPSPVLRQ 920
 QY 961 LHGESPTRLQSKRISDSVSDYDCDGVGVSDYRNGRDLQASTLSVPEQVMSNHC 1020
 DB 921 LHGESPTRLQSKRISDSVSDYDCDGVGVSDYRNGRDLQASTLSVPEQVMSNHC 980
 QY 1021 SPSPGPHRVDTIGRTSMSPSPAPPQNVVEOCHGRTRATGCHYNTISRMDRVRMDHYSS 1080
 DB 981 SPSPGPHRVDTIGRTSMSPSPAPPQNVVEOCHGRTRATGCHYNTISRMDRVRMDHYSS 1040
 QY 1081 DRD-----RDCEAADROPYHESRSTTEORPLAERTTTSRSESR 1118
 DB 1041 ERDSHFTLPRSRHQTSEHHRRDGRDCEAADROPYHESRSTTEORPLAERTTTSRSESR 1100
 QY 1119 PDTNLRSMPSLMTGERSAPPSPALSRSHPTGVSPTSSTPGTGRGRQLPQPKCTL 1178
 DB 1101 ADTNLRSMPSLMTGERSAPPSPALSRSHPTGVSPTSSTPGTGRGRQLPQPKCTL 1160
 QY 1179 ERSAMDIEERNOMKLVYKQVAGSDPLRQDYHESKYRSQMDPHRGADTVTSKSDSDVS 1238
 DB 1161 ER----- 1162
 QY 1239 DVSAVSRSSASRFSSTSYMSQSERPRGRNKISVFTSKMQNRQWVGSKNLTKSTISG 1298
 DB 1163 ----- 1162
 QY 1299 DWSLEKNDQSODTAVGALGTSGKRRSSTGAKWVAIVGLSRKRSASQSQTEGGGKK 1358
 DB 1163 -----GGGKK 1167
 QY 1359 LRSTVORSTETGLAVENRNMTRQASRSTTQGSNNYSSEGNLIPPGVRLASDQPSDFL 1418
 DB 1168 LRSTVORSTETGLAVENRNMTRQASRSTTQGSNNYSSEGNLIPPGVRLASDQPSDFL 1227
 QY 1419 DGLGPAQLVGRQTLATPAWGDIQVGMDDKQQLAVEIIRANGLVYKPKSKTLPATYKVV 1478
 DB 1228 DGLGPAQLVGRQTLATPAWGDIQVGMDDKQQLAVEIIRANGLVYKPKSKTLPATYKVV 1287
 QY 1479 LLDNGVCIKAKTKVARKTLEPLYQQLISFSESQGRVLIQIIVMGDYGMDHKSFMVQAQ 1538
 DB 1288 LLDNGVCIKAKTKVARKTLEPLYQQLISFSESQGRVLIQIIVMGDYGMDHKSFMVQAQ 1347
 QY 1539 ILLDELELSNNVIGWFKLPPSSSLVDPTSAPIPTRASOSSLESSTGSPSBS 1590


```

Db 1116 -----GGGKK 1120
QY 1359 LRSTVORSTTGLAVENRNMWTRQASRESTDGMNSYSSGNLIFFGVRLASQSFDFL 1418
Db 1121 LRSTVORSTTGLAVENRNMWTRQASRESTDGMNSYSSGNLIFFGVRLASQSFDFL 1180
QY 1419 DGLGPAQLVCRQTLATPAMGDIQVGMDDKKGQLEVEIIRARGLVVKGSGNTLPAPYKVY 1478
Db 1181 DGLGPAQLVCRQTLATPAMGDIQVGMDDKKGQLEVEIIRARGLVVKGSGNTLPAPYKVY 1240
QY 1479 LLONGVCIACKTKVARKTLEPLIYQQLLSFEESPOQRVLQIIVWGDYGRMDHKSFGVQAQ 1538
Db 1241 LLONGVCIACKTKVARKTLEPLIYQQLLSFEESPOQRVLQIIVWGDYGRMDHKSFGVQAQ 1300
QY 1539 ILDELELSNMVIGWFKLPPLPPSLVDPPTSAPLTRASQSSLESSTGPSYRS 1590
Db 1301 ILDELELSNMVIGWFKLPPLPPSLVDPPTSAPLTRASQSSLESSTGPSYRS 1352

RESULT 8
Q9JIR9 PRELIMINARY; PRT; 1368 AA.
AC Q9JIR9;
DT 01-OCT-2000 (T-EMBLrel. 15, Created)
DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE RIM2-3A.
GN RIM2-3A.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=20347919; PubMed=10748113;
RA Wang Y., Sugita S., Sudhof T.C.;
RT "The RIM/NIM Family of Neuronal C2 Domain Proteins. Interactions with
RT Rab3 and a new class of Src homology 3 domain proteins.";
RL J. Biol. Chem. 275:20033-20044(2000).
CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC -1- SIMILARITY: CONTAINS 2 C2 DOMAINS.
DR EMBL; AF199325; AAP81647.1; -.
DR HSP; P21707; IRSY.
DR InterPro; IPR000008; C2.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR003315; RPH3A_effector.
DR InterPro; IPR000306; Znf_FYVE.
DR Pfam; PF00168; C2; 2.
DR Pfam; PF00595; PDZ; 1.
DR SMART; SM02318; RPH3A_effector; 1.
DR SMART; SM00239; C2; 2.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS50004; C2_DOMAIN_2; 2.
DR PROSITE; PS00190; CYTOCHROME_C; 1.
DR PROSITE; PS50106; PDZ; 1.
DR PROSITE; PS50178; ZF_FYVE; 1.
SQ SEQUENCE 1368 AA; 155560 MW; CR08E16F0B383A11 CRC64;

Query Match 78.5%; Score 6504; DB 11; Length 1368;
Best Local Similarity 78.8%; Pred. No. 0;
Matches 1283; Conservative 18; Mismatches 29; Indels 298; Gaps 6;

QY 1 MSAPLQPRGPAPTAAOPPPPEMPDLSHLTTEERKIIIVMDQKKEEKESQVLKI 60
Db 1 MSAPLQPRGPAPTAAOPPPPEMPDLSHLTTEERKIIIVMDQKKEEKESQVLK- 59
QY 61 KEHKAQPTQWPPFSGITELVNNVLOPOQKQNEKEPOTKLHQQFEMYEQVKQGEESQ 120
Db 60 -----KLHQQFEMYEQVKQGEESQ 80
QY 121 QOQEQKGADPTGICHTKFPADGCGHNCSCYQTKFCARCGRVSLSNKVMVNCNCRQ 180

```

```

Db 81 QOQEQKGADPTGICHTKFPADGCGHNCSCYQTKFCARCGRVSLSNKVMVNCNCRQ 140
QY 181 QEILTKSGANFNSGNTLQPOQKVPRLGRLEZEAPQEKKAKLHEQPOFGAPGDLSPA 240
Db 141 QEILTKSGANFNSGNTLQPOQKVPRLGRLEZEAPQEKKAKLHEQPOFGAPGDLSPA 200
QY 241 VNGRAHGLTROTITNGSGVHQIADMPDRKSPSVSRDQNRRTYQSEEREDYSQV 300
Db 201 VERGRAHGLTRODSIKNGSGMTHQIADMPDRKSPSVSRDQNRRTYQSEEREDYSQV 260
QY 301 PSDGTMPSPSDYADRSOREPQFTEEPCHLVNTRDNRGRHSHSKYIIVDDSDVSRDEY 360
Db 261 PSDTMTFRSPSDYADRSOREPQFTEEPCHLVNTRDNRGRHSHSKYIIVDDSDVSRDEY 320
QY 361 EKQRESEYQARYSDPMLARYPVQFPEYEQWRIHARVSRARHRRHSDVSLANAELED 420
Db 321 EKQRESEYQARYSDPMLARYPVQFPEYEQWRIHARVSRARHRRHSDVSLANAELED 380
QY 421 SRISILLMDRPSRORSVSERRAAMNORSYSMERTREAOQSSYPORTSNHSPPTPRSP 480
Db 381 SRISILLMDRPSRORSVSERRAAMNORSYSMERTREAOQSSYPORTSNHSPPTPRSP 440
QY 481 IPLDPMRBRADSLAKOHLDPSSAVRKTREKMETMLANDSLSDQSSSVPPPPRPHK 540
Db 441 IPLDPMRBRADSLAKOHLDPSSAVRKTREKMETMLANDSLSDQSSSVPPPPRPHK 500
QY 541 SKKGGMROVSLSSSEKELASTPETVSCDDVLESESEKSDGSKGKRTKESQGVLSDS 600
Db 501 SKKGGMROVSLSSSEKELASTPETVSCDDVLESESEKSDGSKGKRTKESQGVLSDS 545
QY 601 NTRSEKQKRMVYGGHSLFEEDLEWSEPOIKDSGVDTCSTTLNHHSHSDKEHPTWQPSK 660
Db 546 -----YSMLKHASWE-----SSRASPMSLHPTWQPSK 573
QY 661 DGDRLIGRILLNKLKDGSPDSDGAMGLKVVGKMTESGELCAPITVKKGSLADTVG 720
Db 574 DGDRLIGRILLNKLKDGSPDSDGAMGLKVVGKMTESGELCAPITVKKGSLADTVG 633
QY 721 HLRPGDEVLEWNGRLLOGATFEVNIILLESPEQVELVSRPDIPTIPDSHQAOLE 780
Db 634 HLRPGDEVLEWNGRLLOGATFEVNIILLESPEQVELVSRPDIPTIPDSHQAOLE 693
QY 781 SSSSFESQKMDRPSISVTPSPSGMLRDVPQFLSGQLS-----IKLWF 824
Db 694 SSSSFESQKMDRPSISVTPSPSGMLRDVPQFLSGQLSRRITTFVPRVQLKWF 753
QY 825 DKVGHQILVTILGAKDLSPREDCRPNTVYKIYFLPDRSDKKRRTKTKTLEPQWQT 884
Db 754 DKVGHQILVTILGAKDLSPREDCRPNTVYKIYFLPDRSDKKRRTKTKTLEPQWQT 813
QY 885 FIYSPVHRRERFERMLEITLMDQAVRRESEFLEGLILETALLDEPFWYKLOTHDV 944
Db 814 FIYSPVHRRERFERMLEITLMDQAVRRESEFLEGLILETALLDEPFWYKLOTHDV 873
QY 945 SSLPLRPPSPYLPRLQHGHSPTRELQSKRISDSSEVSDICEDGVGVSDYRNRGRDQ 1004
Db 874 SSLPLRPPSPYLPRLQHGHSPTRELQSKRISDSSEVSDICEDGVGVSDYRNRGRDQ 933
QY 1005 SSTLSVPQWSSNHCSPSGSHRVVDVIGRTSNWSPSPAPPPQVNVQCHGRTRATGHYNT 1064
Db 934 SSTLSVPQWSSNHCSPSGSHRVVDVIGRTSNWSPSPAPPPQVNVQCHGRTRATGHYNT 993
QY 1065 ISRMDRHRVMDHYSDSDRD-----RDCEAAQROPVHRSSTEQ 1102
Db 994 ISRMDRHRVMDHYSDSDRD-----RDCEAAQROPVHRSSTEQ 1053
QY 1103 RPLLETTTRSSSRPDTNLRSMPSLMTGRSAPPSPALGRSHPTGVTSVSPSTPCT 1162
Db 1054 RPLLETTTRSSSRPDTNLRSMPSLMTGRSAPPSPALGRSHPTGVTSVSPSTPCT 1113
QY 1163 GRGQLPOLPPKGTFLERSAMDIEERNQMKLNKYQVAGSDPRLBQDTHSKYRSGMDPH 1222

```

```
Db 1114 GRRGRLPQPPKGTGLR----- 1131
Qy 1223 RGADTVSTKSDSDVSDVSAVSTSSASRPSSTSYMSVQSERPRGNRKISVFTSKQNRQ 1282
Db 1132 ----- 1131
Qy 1283 MGVSGKNLTSTISGDMCSLEKNDQSQTAVGALGTGKKRRSSIGAKMVAIVGLSRK 1342
Db 1132 ----- 1131
Qy 1343 SRASQLSQTGGGKKLSTVORSTETGLAVEMRNMTQASRSTDSMNSYSSEGNLI 1402
Db 1132 -----GGGKKLSTVORSTETGLAVEMRNMTQASRSTDSMNSYSSEGNLI 1180
Qy 1403 PGVRLASDSQSPDFLDGIGPAQLVGRQTLATPAMGDIQVGMMDKKGQLEVEIIRARGLV 1462
Db 1181 FGVRLASDSQSPDFLDGIGPAQLVGRQTLATPAMGDIQVGMMDKKGQLEVEIIRARGLV 1240
Qy 1463 VKPGSKTLPAPYVKVYLLDNGVCIARCKTKVARKTLEPLYYQLLSFEESPOQRVLQIIVW 1522
Db 1241 VKPGSKTLPAPYVKVYLLDNGVCIARCKTKVARKTLEPLYYQLLSFEESPOQRVLQIIVW 1300
Qy 1523 GDYGMMDHKSFMGVAQIILLDELELSNMVIGWFKLPPPSLVDPSTAPLTRRASQSLESS 1582
Db 1301 GDYGMMDHKSFMGVAQIILLDELELSNMVIGWFKLPPPSLVDPSTAPLTRRASQSLESS 1360
Qy 1583 TGPSYSRS 1590
Db 1361 TOPSYSRS 1368

RESULT 9
Q8CIX2 PRELIMINARY; PRT; 1363 AA.
AC Q8CIX2;
DT 01-MAR-2003 (TremBLrel. 23, Created)
DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE Synaptic regulatory protein RIM2beta.
GN RIM2BETA.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]_SEQUENCE FROM N.A.
RA Wang Y., Sudhof T.C.;
RT "Genomic Definition of RIM Proteins: Evolutionary Amplification of a
RT Family of Synaptic Regulatory Proteins.";
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF548738; AAN59930.1; -.
SQ SEQUENCE 1363 AA; 154502 MW; C100863603514145 CRC64;

Query Match 75.6%; Score 6261.5; DB 11; Length 1363;
Best Local Similarity 89.8%; Pred. No. 0;
Matches 1230; Conservative 17; Mismatches 24; Indels 99; Gaps 5;

Qy 273 RKSPSVSDQRRRYEQSEREDYQYVPSDCTMPSPSDYADRSQREPQYFEEPHLN 332
Db 41 RKSPSVSDQRRRYEQSEREDYQYVPSDCTMPSPSDYADRSQREPQYFEEPHLN 100
Qy 333 YKDSNRGRHSHKEYIVDDDEYSDRDEYERQREYQARYSDPNLARYPVKQPYEQ 392
Db 101 YKDSNRGRHSHKEYIVDDDEYSDRDEYERQREYQARYSDPNLARYPVKQPYEQ 160
Qy 393 MRIAEVSRARHRRHSDVLANAELEDSRISLLRMDRPSRQSVSERRAAMENQSYSM 452
Db 161 MRIAEVSRARHRRHSDVLANAELEDSRISLLRMDRPSRQSVSERRAAMENQSYSM 220
Qy 453 ERTREAOQSSYPORTNSHSPPTPRSPILDRPDMRADSLRKQHLDPSSAVKTKRE 512
Db 221 ERTREAOQSSYPORTNSHSPPTPRSPILDRPDMRADSLRKQHLDPSSAVKTKRE 280
```

```
Qy 513 KMETMLANDSLSSQSSSVAPPPPPPHKSKKGGKQORQVLSLSSSEELASTPETSCDDVE 572
Db 281 KMETMLANDSLSSQSSSVAPPPPPPHKSKKGGKQORQVLSLSSSEELASTPETSCDDVE 340
Qy 573 LESSEVSEKDSQKGRKTSQGVLSNSTRSRQKRMYYGHSLSLEEDLEWSEPOIKDS 632
Db 341 IESSEVSEKDSQKGRKTSQGVLSNSTRSRQKRMYYGHSLSLEEDLEWSEPOIKDS 363
Qy 633 GVDTCSTTLNEHSHSHKHPTWOPSKOGDLGRILLNKRLLKDGVPDRSGAMGLKV 692
Db 364 -----SSEASPMSLHPVTWOPSKOGDLGRILLNKRLLKDGVPDRSGAMGLKV 413
Qy 693 VGGRTSSEGLCAPIITVKKGSGLADTVGHLRPGDEVLEWNGRLIQATFEVYNILLESK 752
Db 414 VGGRTSSEGLCAPIITVKKGSGLADTVGHLRPGDEVLEWNGRLIQATFEVYNILLESK 473
Qy 753 PEPQVELVSRPIGDIIPRIDSTHAQLESSESSSFESQMDRPSISVTSFMSPGMLRDPQ 812
Db 474 PEPQVELVSRPIGDIIPRIDSTHAQLESSESSSFESQMDRPSISVTSFMSPGMLRDPQ 533
Qy 813 FLSCOLS-----IKLWFKVGHQIIVTILGAKDLPREDGRPNPVYKI 856
Db 534 FLSCOLSQSLSRRTTFFPVQVQIKLWFKVGHQIIVTILGAKDLPREDGRPNPVYKI 593
Qy 857 YFLPDRSDKRRRTKTKVKTLEPKMNQTFIYSPVHRREFRERMLEITLMDQARVEESE 916
Db 594 YFLPDRSDKRRRTKTKVKTLEPKMNQTFIYSPVHRREFRERMLEITLMDQARVEESE 653
Qy 917 FLGEILIELETTALLDDPHWYKLTQHDVSSILPLPPSPYLPFRQYHGESPRRLQSKRI 976
Db 654 FLGEILIELETTALLDDPHWYKLTQHDVSSILPLPPSPYLPFRQYHGESPRRLQSKRI 713
Qy 977 SDSEVSDYDCEDGVGVSDYRNGRDLQSSTLSVPQVMSNHCSPGSPRVDVIGTR 1036
Db 714 SDSEVSDYDCEDGVGVSDYRNGRDLQSSTLSVPQVMSNHCSPGSPRVDVIGTR 773
Qy 1037 SNPSAPPQPNVQEGHGTATGHYNTISMDRHRVMDHYSSDRD----- 1083
Db 774 SNPSAPPQPNVQEGHGTATGHYNTISMDRHRVMDHYSSDRD----- 833
Qy 1084 -----RQCEAADRQPHRSRSTORPILERTTTSRSSSRPDVTLMSRSMSTGR 1134
Db 834 TSEHHRDGRDCEAADRQPHRSRSTORPILERTTTSRSSSRPDVTLMSRSMSTGR 893
Qy 1135 SAPSPALSRSHPRGTGVSQTSPTGTRGRQOLPOLPPKGTGLR----- 1180
Db 894 SAPSPALSRSHPRGTGVSQTSPTGTRGRQOLPOLPPKGTGLR----- 953
Qy 1181 SAMDIEERNRQMLNKYQVAGSDPRLEQDYHSKYRSGWDPHRGADTVSTKSSDSVSDV 1240
Db 954 GAMDIEERNRQMLNKYQVAGSDPRLEQDYHSKYRSGWDPHRGADTVSTKSSDSVSDV 1013
Qy 1241 SAVERTSSASRPSSTSYMSVQSERPRGNRKISVFTSKQSGMVGSKLSTKTSISGDM 1300
Db 1014 SAVERTSSASRPSSTSYMSVQSERPRGNRKISVFTSKQSGMVGSKLSTKTSISGDM 1073
Qy 1301 CSLEKNDQSQDITAVGALGTSGKKRRSIGAKMVAIVGLSRKSASQISQTEGGKKLR 1360
Db 1074 CSLEKNDQSQDITAVGALGTSGKKRRSIGAKMVAIVGLSRKSASQISQTEGGKKLR 1133
Qy 1361 STVQRSTETGLAVEMRNMTQASRSTDSMNSYSSEGNLIFFGVRLASDQSPDFLDG 1420
Db 1134 STVQRSTETGLAVEMRNMTQASRSTDSMNSYSSEGNLIFFGVRLASDQSPDFLDG 1193
Qy 1421 LGPAQLVGRQTLATPAMGDIQVGMMDKKGQLEVEIIRARGLVVKPGSKTLPAPYVKV 1480
Db 1194 LGPAQLVGRQTLATPAMGDIQVGMMDKKGQLEVEIIRARGLVVKPGSKTLPAPYVKV 1253
Qy 1481 DNGVCIAKKTAVKTKLEPLYYQLLSFEESPOQRVLQIIVGDIYGRMDHKSFMGVAOIL 1540
Db 1254 DNGVCIAKKTAVKTKLEPLYYQLLSFEESPOQRVLQIIVGDIYGRMDHKSFMGVAOIL 1313
Qy 1541 LDELELSNMVIGWFKLPPPSLVDPSTAPLTRRASQSLESSSTGPSYSRS 1590
```


Db 1143 QVGMMDKKGQLEVEIIRARGLVVKPGSKTLPAFYTKVYLLONGVCIKKTKKVKARKTLEP 1202

Qy 1501 LYQQLLSFEESPOQGVRLQIIVWGDYGRMDHKSFMGVAQILLDELSNMVIGWFKLPPS 1560

Db 1203 LYQQLLSFEESPOQGVRLQIIVWGDYGRMDHKSFMGVAQILLDELSNMVIGWFKLPPS 1262

Qy 1561 SLVDTSPAPLTRASQSSLESSTGPSYRS 1590

Db 1263 SLVDTPLAPLTRASQSSLESSTGPSYRS 1292

RESULT 11

Q9UQ26 PRELIMINARY; PRT; 1188 AA.

AC Q9UQ26; TREMBLrel. 13, Created

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)

DE Hypothetical protein KIAA0751.

GN KIAA0751.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_Taxid=9606;

RN [1]

SEQUENCE FROM N.A.

RP TISSUE=Brain;

RA MEDLINE=99087487; PubMed=9872452;

RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Miyajima N., Tanaka A.,

RA Kotani H., Nomura N., Ohara O.;

RT Prediction of the coding sequences of unidentified human genes. XI.

RT The complete sequences of 100 new cDNA clones from brain which code

RT for large proteins in vitro.;

RL DNA Res. 5:277-286(1998).

CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.

CC -1- SIMILARITY: CONTAINS 2 C2 DOMAINS.

DR EMBL; AB018294; EAA34471.1; -.

DR HSSP; P21707; IBYN.

DR Genew; HGNC:17283; RIMS2.

DR InterPro; IPR000008; C2.

DR InterPro; IPR001478; PDZ.

DR Pfam; PF00168; C2; 2.

DR Pfam; PF00595; PDZ; 1.

DR SMART; SM00239; C2; 2.

DR SMART; SM00228; PDZ; 1.

DR PROSITE; PS50004; C2 DOMAIN 2; 2.

DR PROSITE; PS50106; PDZ; 1.

KW Hypothetical protein.

SQ SEQUENCE 1188 AA; 135250 MW; DD65DB32288055AF CRC64;

Query Match 68.1%; Score 5638; DB 4; Length 1188;

Best Local Similarity 83.3%; Pred. No. 0;

Matches 1099; Conservative 24; Mismatches 23; Indels 174; Gaps 3;

Qy 273 RKGSVSDQNRVYQSEEREDYQYVPSDGTMPRSQDYADRSQRPQYEPGHLN 332

Db 41 RKGSVSDQNRVYQSEEREDYQYVPSDGTMPRSQDYADRSQRPQYEPGHLN 100

Qy 333 YRDSNRGRHRSKEVIYDDEVEDVESDEYERQRREYQARYRSDPNLARYPVKQPYERQ 392

Db 101 YRDSNRGRHRSKEVIYDDEVEDVESDEYERQRREYQARYRSDPNLARYPVKQPYERQ 160

Qy 393 MEIHAEVSEARHRRHSDVSLANAELEDSRI SLRMDRPSRQSVSERRAAMENQSYSM 452

Db 161 MEIHAEVSEARHRRHSDVSLANAELEDSRI SLRMDRPSRQSVSERRAAMENQSYSM 220

Qy 453 ERTREAOQSSYPTQTSNHSPTPRSPILDRPMDRADSLRKQHLDPSSAVAKTKRE 512

Db 221 ERTREAOQSSYPTQTSNHSPTPRSPILDRPMDRADSLRKQHLDPSSAVAKTKRE 280

Qy 513 KMETMLRNDLSLSDQSESVRPPPPPKHSKKGKMRQVSLSSSEBELASTPEYTCDDVE 572

Db 281 KMETMLRNDLSLSDQSESVRPPPPPKHSKKGKMRQVSLSSSEBELASTPEYTCDDVE 340

Qy 573 LBSVSSEKGDGSKGKTKTSEQVLSDNTRSERQKGMKMYTGHSLSEEDLWSESPQKDS 632

Db 341 LBSVSSEKGDGSKGKTKTSEQVLSDNTRSERQKGMKMYTGHSLSEEDLWSESPQKDS 400

Qy 633 GVDTCSTTLNEEHSKSHKHPVTWQPSKDGRLIGRIILNKLKDGSPVDSGMLGLKV 692

Db 401 GVDTCSTTLNEEHSKSHKHPVTWQPSKDGRLIGRIILNKLKDGSPVDSGMLGLKV 460

Qy 693 VGGNMTESGRLCAPITVKKGLADTVGHLRPGDEVLEWNGRLIQATFEVNIILSK 752

Db 461 VGGNMTESGRLCAPITVKKGLADTVGHLRPGDEVLEWNGRLIQATFEVNIILSK 520

Qy 753 PEPQVELVSRPIGDIPIPDSTHAQLESSESSSFESQMDRPSISVTPSPGMLRDPVQ 812

Db 521 PEPQVELVSRPIGDIPIPDSTHAQLESSESSSFESQMDRPSISVTPSPGMLRDPVQ 580

Qy 813 FLSCQLSILKWFQKVGHLIVTILGANDLPSREDGRPNPVKLYTLPDRSDKDKRTKT 872

Db 581 FLSCQLSILKWFQKVGHLIVTILGANDLPSREDGRPNPVKLYTLPDRSDKDKRTKT 640

Qy 873 VKKTLERKWNQTFYISPVHREFRERMLRITLMDQARVRESEFLEILLKLTALLDD 932

Db 641 VKKTLERKWNQTFYISPVHREFRERMLRITLMDQARVRESEFLEILLKLTALLDD 700

Qy 933 EPHMYKLTQHDVSLPLPRSPYLPFRQLHGESPTRELQSKRISDSSEVDYDCEDGVV 992

Db 701 EPHMYKLTQHDVSLPLPRSPYLPFRQLHGESPTRELQSKRISDSSEVDYDCEDGVV 760

Qy 993 VSDYRNGRDLQSTLSVPQVWSSNHCSPSGSPHRYVDVIGRTSWSPSAPPPQ-RNVEQ 1051

Db 761 VSDYRNGRDLQSTLSVPQVWSSNHCSPSGSPHRYVDVIGRTSWSPSAPPPQ-RNVEQ 820

Qy 1052 CHRGTIA-TGHYNTISRMDRHVRMDHYSDRDRDCEAADRQPHYRHSRSTQRPLEKTT 1110

Db 821 GLGRTTMTGHYNTISRMDRHVRMDHYSDRDRDCEAADRQPHYRHSRSTQRPLEKTT 880

Qy 1111 TRSSSRSDPTNLARSMPSLMTGSAAPPSPALSHPTGVSQVTPSPSTPTGRRGQLP 1170

Db 881 TRSSSRSDPTNLARSMPSLMTGSAAPPSPALSHPTGVSQVTPSPSTPTGRRGQLP 940

Qy 1171 QLPKGTILERSANDIEERNQMKLVKQVAGSDPRLEQDHYHSKYRSGWDPHRGADTVST 1230

Db 941 QLPKGTILERSANDIEERNQMKLVKQVAGSDPRLEQDHYHSKYRSGWDPHRGADTVST 952

Qy 1231 KSSDSDVSDVSAVSRSTSSASRFSSTSYMSVQSERPRGNRKISVFTSKMRQMGVSGNL 1290

Db 953 ----- 952

Qy 1291 TKSTSISGDMCSLEKNDGSDTAVGALGTSGKKRSGSIGANVAIVGLSRKSRASQLS 1350

Db 953 ----- 952

Qy 1351 QTEGKKKLRSTVORSTETGLAVERNWMTQASRSTDSGSMNSVSGENLIIFPGVRLAS 1410

Db 953 -----GKKLRSTVORSTETGLAVERNWMTQASRSTDSGSMNSVSGENLIIFPGVRLAS 1008

Qy 1411 DSQSFDFLDGLGPAQLVGRQTLATPAMGDIQVGMDDKKGQLEVEIIRARGLVKPGSKTL 1470

Db 1009 DSQSFDFLDGLGPAQLVGRQTLATPAMGDIQVGMDDKKGQLEVEIIRARGLVKPGSKTL 1068

Qy 1471 PAPTIVYLLDNGVCIKKTKVARKTLEPLYYQQLLSFERSPOGRVLOIIVWGYGRMDH 1530

Db 1069 PAPTIVYLLDNGVCIKKTKVARKTLEPLYYQQLLSFERSPOGRVLOIIVWGYGRMDH 1128

Qy 1531 KSPMGVAQILLDELSNMVIGWFKLPPSLVDPTSAPIERRASQSSLESSTGPSYRS 1590

Db 1129 KSPMGVAQILLDELSNMVIGWFKLPPSLVDPTSAPIERRASQSSLESSTGPSYRS 1188

RESULT 12

Q9JIR4

ID Q9JIR4

PRELIMINARY;

PRT; 1615 AA.


```

1393 TETGMAERK-MVQPSRESTDGSINSYSEGNLIFPGVRVQDSQFSDFLGLGPAQL 1451
1427 VGRQTLATPANGDIQVGMMDKKQOLEVEIIRAGLVVKGSKTLPAPYKVVLLDNGVCI 1486
1452 VGRQTLATPANGDIQIGMEDKKQOLEVEIRASLTQKQSKSTPAPYKVVLLNGACI 1511
1487 AKKTKVAKTLEPLVQOQLSFESPOGRVQLIIVMGDYGRMDHKSFMGVAQILLDEL 1546
1512 AKKTKIARNTLDPLVQOQLSFESPOGRVQLIIVMGDYGRMDHKSFMGVAQILLDEL 1571
1547 SNNVIGWFKLPFSSLVDPSTASPLTRASQSSLESSTGSPSYRS 1590
1572 SNNVIGWFKLPFSSLVDPSTASPLTRASQSSLESSTGSPSYRS 1615

RESULT 13
035168 PRELIMINARY; PRT; 1553 AA.
AC 035168;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Rab3 effector.
GN RIM.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA MEDLINE=97394473; PubMed=9252191;
RX Wang Y., Okamoto M., Schmitz F., Hofmann K., Sudhof T.C.;
RT "Rim is a putative Rab3 effector in regulating synaptic-vesicle
RT fusion.";
RL Nature 388:593-598 (1997).
CC -|- SIMILARITY: CONTAINS 2 C2 DOMAINS.
CC -|- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
DR EMBL; AF007836; AAB66703.1; -.
DR HSSP; P21707; 1RSY.
DR InterPro; IPR000194; ATPase_a/bcentre.
DR InterPro; IPR000008; C2.
DR InterPro; IPR001360; Glyco_hydro_1.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR003315; RPH3A_effector.
DR InterPro; IPR000306; Znf_FYVE.
DR Pfam; PF00168; C2; 2.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF02318; RPH3A_effector; 1.
DR SMART; SM00239; C2; 2.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS00152; ATPASE_ALPHA_BETA; 1.
DR PROSITE; PS00004; C2_DOMAIN_2; 2.
DR PROSITE; PS00572; GLYCOSYL_HYDROL_F1_1; 1.
DR PROSITE; PS0106; PDZ; 1.
DR PROSITE; PS0178; ZF_FYVE; 1.
DR SEQUENCE 1553 AA; 172964 MW; D6643997E65D03EE CRC64;

Query Match 54.2%; Score 4489.5; DB 11; Length 1553;
Best Local Similarity 57.2%; Pred. NO. 7.2e-273;
Matches 952; Conservative 209; Mismatches 319; Indels 183; Gaps 37;

QY 1 MSAPLGRGRAPTAAPOFPQPEMPDLISLTEREKIILAVMDQKKEEKESQSLVKI 60
DB 1 MSSAVGPRGRPT-----VPPMQELPDLSLTEREENIIMAVMDQKKEEKESQSLVKI 56
QY 61 KSEHTAQP-----TQWPPSGITELVNNVLPQOKOPNKEKPT--KLHQQFEMYKE 110
DB 57 VYRDMAPKPAACKTPRNASQHPPLNIFRCVCPKPSSEGGPERDWRLWQOFESYKE 116
QY 111 QVKMGEESSQOQ--EQKGAPTCGICHTKFAAGCHNCYSYQTFKFCARCGVRSLRNN 169

```

```

DB 117 QVRKI GEARRYQGEHDDADPTCGICHTKFAAGCHNCYSYQTFKFCARCGVRSLRNN 176
QY 170 ----VWVNCNLCRKOELITKSGANFYNSGNTLQOPQKQVGRGLRNE-----EAPQE 218
DB 177 EDKVVWVNCNLCRKOELITKSGANFYNSGNTLQOPQKQVGRGLRNE-----EAPQE 230
QY 219 KXAKLHPOQFOGAPCDLSVPAVEKRAHGLTRQDTIKNGS-----GVKH 263
DB 231 KXAKLHPOQFOGAPCDLSVPAVEKRAHGLTRQDTIKNGS-----GVKH 279
QY 264 QIA----SDMPDRKESVSVDQRRYEOREREDYSQYV--PSDGTWFRSPSYADR 317
DB 280 QKASRSRSEPPRRKKAPGLS--BONGKQOKSRKRVPSKVVQPGEGI-----ADEREK 333
QY 318 SQREPOFYEEPCGLNVRDNRHGHSHSKYIIVDDVEDVESRDYERQREERYQAYRSDP 377
DB 334 ERRETRRLKESQYSD---RPEKRDNGRVAED-----QKQREERYQAYRSDP 381
QY 378 NLARYPVKPPQYEQMRTHAEVSRARHRRHSDVSLANAELEDRISILLMDRPSQRSV 437
DB 382 NLARYPVKPPQYEQMRTHAEVSRARHRRHSDVSLANAELEDRISILLMDRPSQRSV 427
QY 438 SERRAAMENQORSYSMERTREACQSSYQRTSNESPPTPRSPITLDRPMDRADSILNQ 497
DB 428 AERTAGKAPATARVSPSPESPRARAAAQPPTEHGPPPPPPAPGAPRPFEPVPEPLKQ 487
QY 498 HLLDPSSAV--RKTRERKMTMLRNDLSLSDQSESVRPPPPPHKSKKGMQVSLSS 555
DB 488 GRLDPSSAVLLKAKKAKKAKKAKKAKKAKKAKKAKKAKKAKKAKKAKKAKKAKK 547
QY 556 EELASTPEYTSDDVLESESVSEKGSQKGRKTSBQGLSDNTSRTSERQKQKMTGG 615
DB 548 EREGVSTPEYTSDDVLESESVSEKGLD-----YY-- 579
QY 616 HSLIEDLWSEPPQIKDQGVDTCSSTLNEKSHSKHPWTWOPSKDGLRILLNKL 675
DB 580 -----WLDP-----ATWHSRSTPSSSHFVTWQPSKEGDLRILNKLRT 620
QY 676 KQGSVPRDSGAMGLKVVGGKMTSEGRICAPTKVKKGLADTVCHLPQDGLVWNGRL 735
DB 621 ---TWPKESGALLGLKVVGGKMTSEGRICAPTKVKKGLADTVCHLPQDGLVWNGRL 677
QY 736 LGATTFEYVYNIILSKPEPQVELVSRPIGDIPAIPOSTHAQLESSESSSESQMDRPS 795
DB 678 LFGATNEVYVNIILSKPEPQVELVSRPIGDIPAIPOSTHAQLESSESSSESQMDRPS 737
QY 796 ISVTSMPSPGMLRDVPQZLSQGLSKLWFKVGHQILVITLCAKDLPSRQGRPNPYK 855
DB 738 ISVTSMPSPGMLRDVPQZLSQGLSKLWFKVGHQILVITLCAKDLPSRQGRPNPYK 797
QY 856 IYFLPDRSDKMKRTKTKVKKTLPEKWNQTFYSPVHRRERFERMLIETLMDQARVERES 915
DB 798 MYFLPDRSDKMKRTKTKVKKTLPEKWNQTFYSPVHRRERFERMLIETLMDQARVERES 857
QY 916 EFLGELILETALLDDEPHMYKLTQHDVSSLPLPRSPYTLPRRLHGESPTRLQSKR 975
DB 858 EFLGELILETALLDDEPHMYKLTQHDVSSLPLPRSPYTLPRRLHGESPTRLQSKR 917
QY 976 ISDSSEVSDYDCEGVGVYS--DYRNGRDLQSSTLSVPEQVWSSNHCSFSGSPHVDVIG 1033
DB 918 ISDSSEVSDYDCEGVGVYS--DYRNGRDLQSSTLSVPEQVWSSNHCSFSGSPHVDVIG 977
QY 1034 RTRSMSPAPPPQORNEQCH--RGTRATGHTYNTSRM--DRHVRMDHYSDRDRDCEA 1088
DB 978 RPRSELNV--PLQSLDHIHTRSRSTRHDAERSPADSRHSHVESQYISEPDSSELM 1036
QY 1089 ADROPYHRS-----RSTQRFLERTTTRSSSERPTNL---RMSWPSLMT----- 1132
DB 1037 LPRAKGRSASSLWMTSELQPSLDRASASTNCLAFDTLSLSPERHRSKRSKSIQK 1096
QY 1133 GRSAPPSPALSRSHPRTSQVTSSTPGTGRGRLPOLPKK--GTLERSAMDIERNRQ 1191
DB 1097 SRKGTASDA--DRTHRQGSPTQSPADTSFGSRGRGRLPQVPRVSGISQASLIVETRQ 1155

```

QY	1192	M--KLNKYKV--AGSDPRLBQDYHYSKYRSGMWP	PHRGADVTSTKSSDSDYSDVSASVSKTS	1247
DB	1156	MYKVVRFEKQTGSGSQBELDHEQYKSTHKKHQYRS	CNANASAKSDSDYSDVSATSRAS	1215
QY	1248	SASRFSSTVMSVOSRRPCNRKISVPTSQWQNRQMGV	SGKXLTKSTSISGDMCSLEKND	1307
DB	1216	STSRLESTSPWSQSEKPRG--RISSTPQWQRRMTG	TSRAIKSTVSBEIYTLTERND	1273
QY	1308	GSQSDTAVGALGTSGKKRRSIIQAKVVAIVGLSR	KRSASQLSGTQGGKCLASTVORST	1367
DB	1274	GSQSDTAVGTVGAGGKKRSSLKAKVVAIV--SR	SERSTSQLQSTSGHKKLSTVORST	1331
QY	1368	ETGLAVEMRWMTQASRSTDSGMSYSSEGNLIPG	VPLASDSQSFDFGLGPAQLV	1427
DB	1332	ETGMAAEMRK--MYRQPSRSTGDSINSYSEGNL	IPGVRVGPBQF8DFDFGLGPAQLV	1390
QY	1428	GROTLPATPMGDIQVGMMDKKGOLEVEIIRAGLV	VKPKSCTLPAFYKVTLDNGVCIA	1487
DB	1391	GROTLPATPMGDIQIGMEDKKGOLEVEIIRASLT	QKPGSKSTPAPFYKVTLENGACIA	1450
QY	1488	KKTKTVARKTLEPLYOQLLSFFESPOGRVQLIIV	WGDYGRMDHKSFMGVAQIILDELS	1547
DB	1451	KKKTRTARKTLDPLYOQLSLVFDESPQKVLQV	IWVGDIYGRMDHKCFMVAQIILDELDLS	1510
QY	1548	NWVIGWFKLPFPPSSLDVPTSAPLTRASQSSLES	STGSPYSRS	1590
DB	1511	SMWIGWYKLPFPPSSLDVPTLAPLTRASQSSLES	SGPPCIRS	1553

RESULT 14

```

Q8TDY9
ID Q8TDY9 PRELIMINARY; PRT; 1470 AA.
AC Q8TDY9;
AD Q8TDY9;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE RIM long form.
DE RIM long form.
GN NBLA00761.
OS Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RP Aoyama M., Asai K., Shishikura T., Ohira M., Inuzuka H., Morohashi A.,
RT Kato T., Nakagawa A.;
RT "Identification of the alternative form of human RIM." ;
RT Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 2 C2 DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC EMBL; AB051866; BAB87242.1; -.
CC Genew; HGNC:17282; RIMS1.
CC InterPro; IPR000008; C2.
CC InterPro; IPR000345; CytC_heme_bind.
CC InterPro; IPR001360; Glyco_hydro_1.
CC InterPro; IPR001478; PDZ.
CC InterPro; IPR003315; RPH3A_effector.
CC Pfam; PF00168; C2; 2.
CC Pfam; PF00595; PDZ; 1.
CC Pfam; PF02318; RPH3A_effector; 1.
CC SMART; SM00239; C2; 2.
CC SMART; SM00228; PDZ; 1.
CC PROSITE; PS50004; C2_DOMAIN_2; 2.
CC PROSITE; PS00190; CYTOCHROME_C; 1.
CC PROSITE; PS00572; GLYCOSYL_HYDROL_F1_1; 1.
CC PROSITE; PS50106; PDZ; 1.
CC PROSITE; PS50178; ZF_FYVE; 1.
CC SEQUENCE 1470 AA; 164001 MW; 63FD24F4A6E739A CRC64.
SQ

```

Query Match 53.1%; Score 4402; DB 4; Length 1470;
Best Local Similarity 56.4%; Pred. No. 2.1e-267;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 11, 2003, 23:16:12 ; Search time 40 Seconds
(Without alignments)
1869.312 Million cell updates/sec

Title: US-09-617-099B-1
Perfect score: 8285
Sequence: 1 NSAPLGRPRPAPTAASQP.....TRASQSSLESSTGFSYSRS 1590

Scoring table: BLOSUM62

Searched: Gapop 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	637	7.7	5120	1 PCLO_CHICK	Q9pu36 gallus gall
2	555.5	6.7	5147	1 PCLO_HUMAN	Q9y6v0 homo sapien
3	555	6.7	5085	1 PCLO_RAT	Q9jkr6 rattus norv
4	545.5	6.6	5038	1 PCLO_MOUSE	Q9jkr7 mus musculu
5	338	4.1	1021	1 YP77_CABEL	P41885 caenorhabdi
6	307	3.7	1898	1 TRHY_HUMAN	Q07283 homo sapien
7	267.5	3.2	694	1 RP3A_HUMAN	Q9v2j0 homo sapien
8	264	3.2	681	1 RP3A_MOUSE	P47708 mus musculu
9	262.5	3.2	684	1 RP3A_RAT	P47709 rattus norv
10	261.5	3.2	704	1 RP3A_BOVIN	Q08446 bos taurus
11	258.5	3.1	2843	1 APC_HUMAN	P25054 homo sapien
12	258	3.1	1338	1 ACIN_MOUSE	Q9jlx8 mus musculu
13	255.5	3.1	2349	1 TPR_HUMAN	P13270 homo sapien
14	255	3.1	2842	1 APC_RAT	P70478 rattus norv
15	254.5	3.1	2845	1 APC_MOUSE	Q61315 mus musculu
16	254	3.1	2468	1 MAP2_HUMAN	P46821 homo sapien
17	253.5	3.1	1341	1 ACIN_HUMAN	Q9ukv3 homo sapien
18	253	3.1	2440	1 NCRI_HUMAN	P04115 mus musculu
19	248.5	3.0	1453	1 NKCR_MOUSE	P30414 homo sapien
20	244	2.9	1462	1 NKCR_HUMAN	P30414 homo sapien
21	244	2.9	2805	1 MAP2_HUMAN	P78559 homo sapien
22	242	2.9	1861	1 MAP2_RAT	P15146 rattus norv
23	241.5	2.9	1822	1 ZAP3_HUMAN	P49750 homo sapien
24	241	2.9	1549	1 TRHY_SHEEP	P22793 ovis aries
25	239.5	2.9	2476	1 ATRX_MOUSE	Q61687 mus musculu
26	233.5	2.8	1407	1 TRHY_RABIT	P37709 cryptotagus
27	232	2.8	1983	1 TP20_MOUSE	Q9epq8 mus musculu
28	230.5	2.8	2453	1 NCRI_MOUSE	P60974 mus musculu
29	225	2.7	1359	1 ATRX_CABEL	Q9u7e0 caenorhabdi
30	224.5	2.7	4377	1 ANK3_HUMAN	Q12955 homo sapien
31	223	2.7	2142	1 BAT2_HUMAN	P48634 homo sapien
32	221.5	2.7	1828	1 MAP2_MOUSE	P20357 mus musculu
33	221	2.7	2492	1 ATRX_HUMAN	P46100 homo sapien

ALIGNMENTS

RESULT 1

ID	PCLO_CHICK	STANDARD;	PRT;	5120 AA.
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Piccolo protein (Aczonin) (fragment).			
GN	PCLO OR ACZ.			
OS	Gallus gallus (Chicken).			
OC	Rukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;			
OC	Gallus.			
OX	NCBI_TaxID=9031;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Brain;			
EX	MEDLINE=99439764; PubMed=10508862;			
RA	Wang X., Kibechull M., Laue M.W., Lichte B., Petrasch-Parwez E.,			
RA	Kilimann M.W.;			
RT	"Aczonin, a 550-kd putative scaffolding protein of presynaptic active			
RT	zones, shares homology regions with rim and bassoon and binds			
RT	profilin.";			
RL	J. Cell Biol. 147:151-162(1999).			
CC	- FUNCTION: May act as a scaffolding protein involved in the			
CC	organization of synaptic active zones and in synaptic vesicle			
CC	trafficking (By similarity).			
CC	- SUBUNIT: Interacts with Rabac1/Pral and profilin (By similarity).			
CC	- SUBCELLULAR LOCATION: Concentrated at the presynaptic side of			
CC	synaptic junctions (By similarity).			
CC	- DOMAIN: C2 domain 1 is involved in binding calcium and			
CC	phospholipids. Calcium binds with low affinity but with high			
CC	specificity and induces a large conformational change.			
CC	- SIMILARITY: Contains 2 C2 domains.			
CC	- SIMILARITY: Contains 1 PDZ/DHR domain.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL; Y19187; CAB60725.1; --			
CC	HSSP; P04410; IJ25.			
DR	GO; GO:0045202; C:synaptic junction; ISS.			
DR	GO; GO:0005509; F:calcium ion binding activity; ISS.			
DR	GO; GO:0005544; F:calcium-dependent phospholipid binding acti. . .; ISS.			
DR	GO; GO:0005522; F:profilin binding activity; ISS.			
DR	GO; GO:0007010; P:cytoskeleton organization and biogenesis; ISS.			
DR	GO; GO:0016080; P:synaptic vesicle targeting; ISS.			
DR	InterPro; IPR000008; C2.			
DR	InterPro; IPR001478; PDZ.			
DR	InterPro; IPR001565; Synaptotagmin.			
DR	Pfam; PF00168; C2; 2.			

Q9udy2 homo sapien
Q99hb2 mus musculu
Q02952 homo sapien
P34926 rattus norv
P15205 rattus norv
Q8m881 drosophila
Q92340 rattus norv
P27715 caenorhabdi
Q13428 homo sapien
Q92817 homo sapien
Q95168 canis fami
Q920u1 mus musculu

34 220.5 2.7 1190 1 Z02_HUMAN
35 220 2.7 1333 1 PAD3_MOUSE
36 218 2.6 1781 1 AKL2_HUMAN
37 217.5 2.6 2774 1 MAP2_RAT
38 215.5 2.6 2459 1 MAP2_RAT
39 215.5 2.6 2779 1 LVA_DROME
40 215 2.6 1337 1 PAD3_RAT
41 214.5 2.6 1813 1 UN13_CABEL
42 214 2.6 1411 1 TCOF_HUMAN
43 213 2.6 2033 1 EVPL_HUMAN
44 211 2.5 1174 1 Z02_CANFA
45 209 2.5 1167 1 Z02_MOUSE

[illegible]

VSP_003926, VSP_003927;
Note-No experimental confirmation available;
-1- DOMAIN: C2 domain 1 is involved in binding calcium and phospholipids. Calcium binds with low affinity but with high specificity and induces a large conformational change.
-1- SIMILARITY: Contains 2 C2 domains.
-1- SIMILARITY: Contains 1 PDZ/DHR domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

EMBL; Y19188; CAB60727.1; --
EMBL; AC004903; AAD20936.1; --
EMBL; AC004866; AAD21789.1; --
EMBL; AB011131; BAA25485.1; --
EMBL; BC001304; AAH01304.1; --
EMBL; AC004082; AAB97937.1; --
PIR; T00634; T00634.
HSSP; P04410; LA2S.
Genew; HGNC:13406; PCLO.
MIM; 604918; -- C:cytoskeleton; NAS.
GO; GO:0005856; C:synaptic junction; ISS.
GO; GO:0045202; C:synaptic ion binding activity; ISS.
GO; GO:0005509; F:calcium ion binding activity; ISS.
GO; GO:0005544; F:calcium-dependent phospholipid binding acti...; ISS.
GO; GO:0005522; F:profilin binding activity; ISS.
GO; GO:0007010; P:cytoskeleton organization and biogenesis; ISS.
GO; GO:0016080; P:synaptic vesicle targeting; ISS.
InterPro; IPR000008; C2.
InterPro; IPR001565; Synaptotagmin.
PRINTS; PR00360; C2DOMAIN.
FRINTS; FR00399; SYNAPTOTAGMN.
SMART; SM00239; C2; 2.
PROSITE; PS00499; C2_DOMAIN_1; 1.
PROSITE; PS00004; C2_DOMAIN_2; 2.
Calcium/phospholipid-binding; Zinc; Metal-binding; Zinc-finger;
Repeat; Alternative splicing.
NON TER 1
DOMAIN 400 465

10 X 10 AA TANDEM APPROXIMATE REPEATS OF
P-A-K-P-Q-P-Q-Q-P-X.
C4-TYPE (POTENTIAL).
C4-TYPE (POTENTIAL).
POLY-PRO.
PDZ.
C2 DOMAIN 1.
C2 DOMAIN 2.
S -> SGNGLGIRIVGGKEIPGHSGISGAYAKILPGSGAE
QTGKLMKG (in isoform 2).
/FTID-VSP_003923.
K -> KPTDQGVSHVTHGTIQ (in isoform 2).
/FTID-VSP_003924.
G -> GQNVNQAS (in isoform 2).
/FTID-VSP_003925.
TAHKS -> SKRRK (in isoform 2).
/FTID-VSP_003926.
Missing (in isoform 2).
/FTID-VSP_003927.
SEQUENCE 5147 AA; 563537 MW; CD5D84990498CD3C CRC64;

VARSPLIC 4534 4534
VARSPLIC 4576 4576
VARSPLIC 4757 4761
VARSPLIC 4762 5147
SEQUENCE 5147 AA; 563537 MW; CD5D84990498CD3C CRC64;

Query Match 6.7%; Score 535.5; DB 1; Length 5147;
Best Local Similarity 21.8%; Fred. No. 2.1e-18;
Matches 308; Conservative 208; Mismatches 488; Indels 407; Gaps 63;

316 RSRSREPOYTEEPGHLYNRDSNRCHRESKEYIYVDDEVERSEDTYE-----KPRHEETQ- 370
3981 RRSQVTDF-----LAPQSRRLLH-----SYVAEED-FMEDPYHKLKHQTKOEPR 4039

QY 371 -----ARVSDPNLARYPVKPPQYEQ--WRHAHV---SRARHER 407
 Db 4030 GTESDLHLAGLHYHADTSYRHPFKSEKYSRLTLEKQAAKALPAILYKQSKHKKS 4089
 QY 408 HSDVSLAN-AELEDKRIILLRMDRPRQRSS-----VSERRAAMENQSYSMERT--- 455
 Db 4090 LIDPMKSPFIOERD--LEPDYSYMTSSISSIGISSARALLQDDITFOLRKNITDQ 4147
 QY 456 -----REA-QQSSYPORTSNHSPPTPRSPIT-PLDRPDMRADDL 494
 Db 4148 QKFMSSILGTGLTGLTIRTSALQDEADKPYSGSGSRSPSPSPSVYGLDL-----SI 4200
 QY 495 RQCHLDPSAVKTKREKMEWMLNDLSLSPQSESVPPPPRPHKSKGKGKGRQVSLSS 554
 Db 4201 KR-----DSSSSSLRLKAQEAEL-----DVSFSSASAKTKPTSLPISQSRGRIPVAQNS 4253
 QY 555 SEERLASTPEYTSDDVLESSEVSEKGDQKKEKTSQGVLSNTRSRQKRMYYG 614
 Db 4254 EEE-----SPLSPVQPMWARAAAGPLPIGADTDQ-----FGS 4289
 QY 615 GHSLREDLEWSPQIKDGDVTCSTTLNE-EHSHSD-----KHYVTWQSPKDGRLI 666
 Db 4290 SHSLPEVQCHMEESKRTGYDRDIAFIMDDFQAMSDSEAYHLRREETDMFKPRESLE 4349
 QY 667 GRILLKELKDGSPRDSGAMGLKAVGKMTESGELCAPI-----TKVKGSLADTGH 721
 Db 4350 NGHGLDRKLPRLVHRSRPLSQHQBQII-----QMNKTHYIFPHARIKITRDSKDTYSM 4405
 QY 722 LRPQDEVLEWNGRLQCATFREVMIILLESKEPOVELVVRPIGDIPIRDPSTHAQLES 781
 Db 4406 -----QVLEWNGIPITSKTYEVSQII---SQSGEATICVRLDMLWLSSENSQHLLEH 4458
 QY 782 SSSSFESQK-----MDRPSISVTSPM 802
 Db 4459 PPKAVDKAKSPGDPKQAAELQKVSLOQSPVLSSVVEKSHVSGPTSAGSSVSPSG 4518
 QY 803 SPGLMDVPQFSLQSLKLMFKDKVGHQILVILGAKDLPSREDGRPNPYKLYFLDR 862
 Db 4519 QPGS-PSVSKKKGSGSKLQIYND-LGN-LIITHILOARNLVPRDNGNGSDPFVYLLPGR 4575
 QY 863 SDKNRRRTKVKTLPEKWNQTFIYSPVHRFRFRMLBITLMDQARVRESESEFLGIL 922
 Db 4576 GAEYKERTKHQKSLNPNWQIVYKISMEQKKKLTLEVTDYDRP--SSNDFLGEVL 4633
 QY 923 IEL-ETALLDDPHWKL--QTHDVSSLPPLRPSYLPRLQHLGSS-PTRELQSRKISD 978
 Db 4634 IDLSSTSHLDNTPRWYPLKEQTESID-----HGKSHSSQSSQSPKPS- 4676
 QY 979 SEVSDYDCEDGVGVSDYTHNGRDLQSSLTSLVPEQVMSSNHCSPSGSPHRVDVIGRTSM 1038
 Db 4677 -----VIKSRSHGIFPPDPKDMQVP--TIKSHSSPGSGSKSSB--GHLRSH 4719
 QY 1039 SPSPAPPQNVQCHRGTRATGHYNTISMRDRHVMDHYSDDRDCEAADRQPHYRSR 1098
 Db 4720 GFSRQSKTSVQTH-----LED--AGAAIAAAEAQQ----- 4751
 QY 1099 STEQRPLERTTTRSSSRPPTNLMRSPSL-MTGRSAPPSPALSRHPTFGSVQTPS 1157
 Db 4752 -----LRIOPTAKSGQS-----NHARKQHRHSIAG 4777
 QY 1158 STPGTGRGRQLPOLPPKGLTSLERSAMDIERRQMKLKYQVAGSDPLEQDHYSKYRS 1217
 Db 4778 VLPQRTQSDNLPP-PANGNQDS-----QIALRKVMSDGPVKP-----E 4816
 QY 1218 GMDP--HRGAD--TVSTKSDSDVSDVSAVRSSTSSASRFSSTVMSVQSERPRGNKISVF 1274
 Db 4817 GAKPMHRAESVSTGSGSGSGSGYSDSGSSSTAGETMLFI-----PRIGK----- 4867
 QY 1275 TSKMQRQMGVSGKMLKTSISGDMCSLEKNDGSDQSTAVGALGTSGKGRSS----- 1328
 Db 4868 -----MGQNGQBPVK-----QPGVGVLADTEARGQREPKPGQAC 4902
 QY 1329 -IGAKMVAIVGLSRKRSASQLSQTEGGQKLRSTV-----QRSTETGLAVENRNM 1379

Db 4903 FIGAR-----NWKSIQMPHEIYPTDKGKYGSLFPFSLNCCOIKT-----NW- 4949
 QY 1380 TRQARSTDCSMNSYSSEKML-IFPGV---KLASDSQF-----SDFLDGLG 1422
 Db 4950 -----KPVDS-----EAQADLELFLNYTRILPPSYLLQKTKALQTYERSDIRSGNG 4998
 QY 1423 PAQIVRGQTATPANGDIQVM---MDKKG-QLKVEIIRARGLVVKGPSKT-LPAPYVKV 1477
 Db 4999 FTINATTCFTTSVWGEIKALKKEMKTDGBQLIVEILQCCNITVYKPSPDHLPOLYVKI 5058
 QY 1478 YLLONGV--CIAKKTKVAKTLEPLVQQLSFEESQQRVLOIIVWGDYGMHKSFMG 1535
 Db 5059 YVMISTQKRYKIKKTRVCRHREPSFNETRPSLSPAGHSLOILLFSNOGKFMKTKLLIG 5118
 QY 1536 VAQILLDELSNMVIGNFKLFPSPSLVDPT 1566
 Db 5119 EACIWLNDKRLKRVNVHKL-----LVSPT 5144

RESULT 3
 PCLO RAT
 ID PCLO RAT STANDARD; PRG; 5085 AA.
 AC Q9JXS6; OSJTLI;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Piccolo protein (Multidomain presynaptic cytomatrix protein).
 GN PCLO.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 2), AND INTERACTION WITH RABAC1.
 RP MEDLINE=20170257; PubMed=10707984;
 RA Fenster S.D., Chung W.J., Zhai R., Cases-Langhoff C., Voss B.,
 RA Garner A.M., Kaempf U., Kindler S., Gundelfinger E.D., Garner C.C.;
 RT "Piccolo, a presynaptic zinc finger protein structurally related to
 RT bassoon.",
 RT Neuron 25:203-214 (2000).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RA Fenster S.D., Cases-Langhoff C., Gundelfinger E.D., Garner C.C.;
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBSJ databases.
 RN [3]
 RP CALCIUM-BINDING ACTIVITY, AND MUTAGENESIS OF ASP-4668; ASP-4674;
 RP VAL-4688; MET-4689; VAL-4690; SER-4691; GLN-4692; ASN-4693 AND
 RP ALA-4694.
 RX MEDLINE=21181819; PubMed=11285225;
 RA Gerber S.H., Garcia J., Rizo J., Suedhof T.C.;
 RT "An unusual C(2)-domain in the active-zone protein piccolo:
 RT implications for Ca(2+) regulation of neurotransmitter release";
 RL EMBO J. 20:1605-1615 (2001).
 CC -!- FUNCTION: May act as a scaffolding protein involved in the
 CC organization of synaptic active zones and in synaptic vesicle
 CC trafficking (By similarity).
 CC -!- SUBUNIT: Interacts with Rabac1/Pral and profilin.
 CC -!- SUBCELLULAR LOCATION: Concentrated at presynaptic side of synaptic
 CC junctions.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing, Named isoforms=2;
 CC Name=1;
 CC Name=2;
 CC Name=3;
 CC IsoId=Q9JXS6-1; Sequence=Displayed;
 CC IsoId=Q9JXS6-2; Sequence=VSP_003930, VSP_003931;
 CC -!- DOMAIN: C2 domain 1 is involved in binding calcium and
 CC phospholipids. Calcium binds with low affinity but with high
 CC specificity and induces a large conformational change.
 CC -!- SIMILARITY: Contains 2 C2 domains.
 CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

QY 1397 SEGNLFFGVRLASDSQSFDFLDGLGPAQLVGRQTLATPAMGDIQVGMKKGQLEVEHII 1456
 Db 4968 -----OLIVETL 4974
 QY 1457 RARGLVVKGSKT-LPAPYVVKVYLLDNGV--CIAKKTKVARKTLEPLVQQLSFEESPO 1513
 Db 4975 QCRNITYKFKSPDHLVLDLVKLVINISQKRVKIKKTVCHDRFSPNFWFRSLSPA 5034
 QY 1514 GRVLQIIVGWYGRMDHKSFMGVAQILLDELSNNVIGWFKLFPSSIVDPT 1566
 Db 5035 GHSLSQILLFSGNGKFMKTKLIGBACIWLDKVLRKRVNWHKL-----LVSPT 5082

RESULT 4

PCLO_MOUSE
 ID PCLO_MOUSE STANDARD; PRT: 5038 AA.
 AC Q9QYX7; Q9QYX6; Q9QZJ0;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Piccolo protein (presynaptic cytomatrix protein) (Aczonin) (Brain-
 derived HLMN protein).
 GN PCLO OR ACZ.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A., SUBCELLULAR LOCATION, ALTERNATIVE SPLICING,
 RP TISSUE SPECIFICITY, AND INTERACTION WITH PROFILIN.
 RC TISSUE=Brain;
 RX MEDLINE=99439764; PubMed=10508862;
 RA Wang X., Kibschull M., Laue M.M., Lichte B., Petrasch-Parwez E.,
 RA Killmann M.W.;
 RT "Aczonin, a 550-kd putative scaffolding protein of presynaptic active
 RT zones, shares homology regions with rim and bassoon and binds
 RT profilin".
 RL J. Cell Biol. 147:151-162 (1999).
 RN [2]
 RP REVISIONS.
 RC TISSUE=Brain;
 RA Killmann M.;
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 4502-4682 FROM N.A.
 RC TISSUE=Brain;
 RA Huang W., Jin W., Huang C., Chen B., Zhang J., Ju G.;
 RL Submitted (OCT-1999) to the EMBL/GenBank/DDAJ databases.
 CC -!- FUNCTION: May act as a scaffolding protein involved in the
 CC organization of synaptic active zones and in synaptic vesicle
 CC trafficking.
 CC -!- SUBUNIT: Interacts with Rabac1/Pral and profilin.
 CC -!- SUBCELLULAR LOCATION: Concentrated at the presynaptic side of
 CC synaptic junctions.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=Q9QYX7-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q9QYX7-2; Sequence=VSP 003928, VSP 003929;
 CC -!- TISSUE SPECIFICITY: Highly expressed in brain. Low levels found in
 CC stomach. Not detected in other tissues analyzed including adrenal
 CC gland, testis and pancreas.
 CC -!- DOMAIN: C2 domain 1 is involved in binding calcium and
 CC phospholipids. Calcium binds with low affinity but with high
 CC specificity and induces a large conformational change.
 CC -!- SIMILARITY: Contains 2 C2 domains.
 CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; Y19185; CAB60731.2; -;
 DR EMBL; Y19186; CAB60732.2; -;
 DR EMBL; AF181269; AAD55786.2; -;
 DR HSSP; P04410; IAJ5.
 DR MGD; MGI:1349390; Pclo.
 DR GO; GO:0045202; C:synaptic junction; IDA.
 DR GO; GO:0005509; F:calcium ion binding activity; ISS.
 DR GO; GO:0005544; F:calcium-dependent phospholipid binding acti. . .; ISS.
 DR GO; GO:0005522; F:profilin binding activity; IDA.
 DR GO; GO:0007010; P:cytoskeleton organization and biogenesis; IDA.
 DR GO; GO:0016080; P:synaptic vesicle targeting; NAS.
 DR InterPro; IPR000008; C2.
 DR InterPro; IPR001478; PDZ.
 DR Pfam; PF00168; C2; 2.
 DR Pfam; PF00595; PDZ; 1.
 DR SMART; SM00239; C2; 2.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS00499; C2 DOMAIN 1; 1.
 DR PROSITE; PS00004; C2 DOMAIN 2; 2.
 DR PROSITE; PS00106; PDZ; 1.
 KW Calcium/phospholipid-binding; Metal-binding; Zinc; Zinc-finger;
 KW Repeat; Alternative splicing.
 FT DOMAIN 371 470
 FT 10 X 10 AA TANDEM APPROXIMATE REPEATS OF
 FT P-A-K-P-Q-P-Q-P-X.
 FT C4-TYPE (POTENTIAL).
 FT C4-TYPE (POTENTIAL).
 FT POLY-PRO.
 FT PDZ.
 FT C2 DOMAIN 1.
 FT C2 DOMAIN 2.
 FT TKPTN -> SKRKK (in isoform 2).
 FT /FTID=VSP 003928.
 FT Missing (in isoform 2).
 FT /FTID=VSP 003929.
 FT VARSPLIC 4834 5038
 FT VARSPLIC 4834 5038
 FT SEQUENCE 5038 AA; 547600 MW; DADA460CF3B40888 CRC64;

Query Match 6.6%; Score 545.5; DB 1; Length 5038;
 Best Local Similarity 20.6%; Pred. No. 6.2e-18;
 Matches 295; Conservative 192; Mismatches 384; Indels 561; Gaps 57;
 QY 316 RRSQEPQFTEPGHLNYSNRGRHSKSYIYDDEVEDVESRDYTE-----RQRREERYQ- 370
 Db 3984 RRSQEVTF-----LAPLQTSRLH-----SYVKAED-FMEDPYELKLLKHQKQEPFR 4032
 QY 371 -----ARTSDPNLARYPVKQPYTEQ--MRIHAEV-----SRARHRR 407
 Db 4033 QTESLDHLAAGLSHYTHADTSYRHFPEKSEKYSIRLTLEKQAKQLPAALYKQSKHKKA 4092
 QY 408 HSDVSLAN-AELEDRIILLRMDRPSRORS-----VSERRAAMENQRSYSMERT----- 455
 Db 4093 LIDPKMSKFSPIQSRD--LEPDITYLSSTSSIGSSARLLQDDITPLGRNITDQ 4150
 QY 456 -----REA-CQSSYPORTNSHPPTPRSP1-PLDPFMRADSL 494
 Db 4151 QKFWSSILGSLGTLGWIRTSALQDEADKPYSSGSRSPSPSPSVYCLDI-----SI 4203
 QY 495 RQCHLDPSAVRTKREKMTMLRNLSLSDQSSEVPPPPPHKHKKGKGMQVSLSS 554
 Db 4204 KR-----DSSSSSLKKAQAEAL---DVSGHSSSSARTKFTSIPFGSRGRIPVAQNS 4256
 QY 555 SESEELAS-----TPETSCDDVLESGSVSKGDSQNGKRT----- 591
 Db 4257 REESPLSPVQOPMGMAAAGPLPIADTKDQFGSSHSLEPVOHREESRTGYRDI 4316
 QY 592 -----SEQVLSDNSTRSEROKRWY-----GHSLE-----EOLWSEPOIK-DS 632
 Db 4317 AFIMDDQHAMSDSEAYHLREEDTDKDPRESLENGHGLDRKLPRLVHSRPLSQHOE 4376
 QY 633 GVDTCSTTLNEEHSNDKHPVTWQPKDGRLLTGRILLNKLNDKGSVPRDSGAMGLKV 692

```

Db      4377 QILQMGKTHYIFPPA-----RIKTRDSKHTV---SGNLGIRI 4415
QY      693 VGGKMT--BSGRCAPIITVKKGSLADTVGHLPFGDEVLEFWNRLQAGAFEEVYNILIE 750
Db      4416 VGGKEIPGHSIGAVIAKILPGSARHSKLEGMQVLEWNGIPITSTYEEVQIINQ 4475
QY      751 SKPEPQV-----ELVSRPIGDIPIPDSTHAQ--LESSSSSSESQW-- 791
Db      4476 QSGEAEICVRLDLNMLSDSENPOHLEHPE-----PKVVDKAKSPGVDPKLAARLQVSL 4531
QY      792 -----DRPSISVTSMSPG-----MLRDVPQ 813
Db      4532 QQSPLVSSVVEGAHAGSPSAGSSVPSQPGSPSVSKKKGSKPTDVSKTASHP 4591
QY      814 LSGQLSIXLWFKVGHQLIWTILGANDLPSREDGRPNPVYKIFLPDR-----SD 864
Db      4592 ITGEIQLQINYD--LGN--LIHILQARNLAVPRDNNGYSDFPVKYLAPGRQVWVQNASV 4649
QY      865 KNRKRTKTKTKLEPKWNTPIYFVRRERFERMLEITLMDQARVRESRFLGHLIE 924
Db      4650 EYKRTKYVKSINPEWNTVIYKSIEMQLMKKTLEVTVMYDRF--SSNDFLGEVLID 4707
QY      925 L-ETALLDDEPHWIKL--QTHDVSSLPLRPSYLPRLRQLHGES--PTRELQSKRISDSE 980
Db      4708 LSTSHLNTPRYPLKEQTESIE-----HGKSHSSQSQSPKPS----- 4748
QY      981 VSDYDCEGQGVVSVYRHNRLQSLTSLVPSQVMSNHCSPSGSPHRYVDVIGRTSMSP 1040
Db      4749 -----VIKSRSHGIFPDPSKDMQVP--TIKSHSSPQSSKSSS--GHLRSHCP 4793
QY      1041 SAPPQPNVQOHRGTRATGHYNTISRMDHRVMDHYSSDRDRDCEAADRQFYHRSRT 1100
Db      4794 SRSQKTSVAOTH-----LED--AGAAIAAAEAQVQ----- 4823
QY      1101 EQRPLLEKTRTSRSPRDTNLMRSMPSLMTGRSAPPSPALSRSHPRGTSVQTSPTSP 1160
Db      4824 -----LRIQP-----TKPTN-- 4833
QY      1161 GTGRGRQLPOLPPKGTLLERSAMDIEERNOMKLNKIKQVAGSDPRLEQYHYSKYSRWD 1220
Db      4834 ----- 4833
QY      1221 PHRGADT-VSTKSSDSVSDVSAVARTSSASRPSSTVMSVQSRPRGNRKISVFTSKWQ 1279
Db      4834 -HRPATSIVSTSGSSVG-----SGYSVDSE----- 4859
QY      1280 NFGMGVGNLTKSTSISDMCSLEKNDSQSDTAVGALGTSKK--RESSIGAKWALV 1337
Db      4860 -----GSSCVAFENLL-----PIPRIGQMGQDPVQKPGMGA----- 4894
QY      1338 GLSKRSASQLSOTEGGKKLSTVORSTETGLAVEMRNMTQASRSTGSMNSYSS 1397
Db      4895 -----ADTEA-----KTQVMGEIKLAKKMK-----TDEE----- 4920
QY      1398 EGNLIPPGVRLASDSQSFDLGLGPAQLVGRQTLATPAMGDIQVGMMDKKGQLSVIR 1457
Db      4921 -----QLIVEILQ 4928
QY      1458 ARGLVVPGSKT-LPAPYVKKVLLNGV--CIAKKTKVARTNLEPLVQQLLSPEESQ 1514
Db      4929 CENITTKSPDLPLDLYKIVINIAQKVIKKTIVCRHREPSRNETRFLSPAG 4988
QY      1515 RVLIIVMGDGRMDHKSFMGVAQIILLDELSNNVIGNFKLFPSPSLVDPT 1566
Db      4989 HSLQILLFSGNGKFKMKTILGEACTWLDKVDLRKRVNWKL-----LVSPT 5035

```

RESULT 5

YPT7 CAEL

ID YPT7 CAEL

AC P41685; STANDARD; PRT; 1021 AA.

DT 01-NOV-1995 (Rel. 32, Created)

```

DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 111.7 kDa protein F37A4.7 in chromosome III.
GN F37A4.7.
OS Caenorhabditis elegans.
OC Rukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Feloderinae; Caenorhabditis.
NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Fulton L., Waterston R.;
RL Submitted (FEB-1994) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Contains 2 C2 domains.
CC -!- SIMILARITY: Contains 1 FYVE-type zinc finger.
CC -!- SIMILARITY: SOME, TO YEAST YNL087W.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U00032; AAA50635.1; -.
DR PIR; S44644; S44644.
DR HSSP; P47709; 1ZBD.
DR WormPep; F37A4.7; CR00705.
DR InterPro; IPR000008; C2.
DR InterPro; IPR002149; LRI.
DR InterPro; IPR003315; RPH3A effector.
DR InterPro; IPR001565; SynapToTagmin.
DR InterPro; IPR00306; Znf_FYVE.
DR Pfam; PF0168; C2; 2.
DR Pfam; PF02318; RPH3A effector; 1.
DR PRINTS; PR00360; C2DOMAIN.
DR PRINTS; PR00399; SYNAPTOTAGMIN.
DR SMART; SM0239; C2; 2.
DR PROSITE; PS00499; C2 DOMAIN 1; 2.
DR PROSITE; PS50004; C2 DOMAIN 2; 2.
DR PROSITE; PS50178; ZF_FYVE; 1.
KW Hypothetical protein; Repeat; Zinc-finger.
FT ZN FING 168 233
FT DOMAIN 758 847
FT DOMAIN 898 987
FT SEQUENCE 1021 AA; 111749 MW; 947C38B7C37BA462 CRC64;
SQ
Query Match 4.1%; Score 338; DB 1; Length 1021;
Best Local Similarity 20.9%; Pred. No. 6.1e-09;
Matches 197; Conservative 139; Mismatches 342; Indels 266; Gaps 43;
QY 90 KQPEKEPQTKLHQPFMYKQVKM---GREGSQOQOQK-----GDAP 130
Db 114 RSTNSKAQTSITAAE--QEHQKVLAKAESKESKQRIKRMVDRLEKMRERATNGV 171
QY 131 T-CGICHTKTF---ADGCGHCNVCYQTKFCAR--CGGR---VSLRSNKV--MWVNLCK 179
Db 172 THCLLCH-TEGLLASKSYAAMCVDCRYVCQRCNGVETTVDTVGTQVETVFLCKICSE 230
QY 180 QOEIL-TTSGAWFYNSGNTLQPDQKVPRLG---RNEEAFQEKKAKLHQPQPGAPG- 234
Db 231 ARVLMKKSQAWFYKE-MPEFQREDDRLPYVTVVTTNGTLFNASSA-----ATPLSGTFFG 285
QY 235 ----DLSPAVE-----KRAEGLTRQDTIKNGSVKQIADWPSDRFS----- 276
Db 286 AGPQMTMPSTSSCOMTTPKNASPGVCSQCLQNGCP-----TSPLNATKRWTHGCGIE 341
QY 277 -PSVSRQNRRYEQSEREDYSQIVP-----SDGTMPSPSDYADRRSQRPQTFEPOHL 331
Db 342 PFRSSDQESFVQGVPRRLNNTKTPVGSATTSAPPPTSTTTFTTSRE-----A 392
QY 332 NYRDSNRGRHRSKEYIVDDDEVSRDEYERQREBEYQARYSDPNLRYFVKPQFVEE 391

```

Db 393 NMERSRTHANRKYSDDDSSPSPBTRGTS-----PHSL----- 434
 Qy 392 QMRHAYSRARHRRHSDVSLANAELEDRSILRMDRSPQRVSERRAMENQRSYS 451
 Db 435 -----ATPSSVAHDTC-DTSDFDADTSIDGVSQSDHSPQOS-----GLTCS 478
 Qy 452 MERTREAOQSSYPQRTSNHS-PPTPRSPPLDRPDMRRADSLRKQHLDPSSAVRTK 510
 Db 479 SSSLTFLQQAASH-----DHSGGCTPR-----ISNPD-----RTTS 511
 Qy 511 REKMETMRLNDSLDQSESVRPPPHKSKKGKQMGQVSLSSSEELASTPEVTSDD 570
 Db 512 RV-----AQASAGTSLVTPPPP-----ISSRTSPDNCNSPLNV----- 545
 Qy 571 VELESSEYSEKDSQKGRKT-SEQVLSDSNTRSEROKRMYGCHSLEEDLESEPOI 629
 Db 546 -----MEHKSASASTASSGGRNRVGSAPVLNNHMMNNO-----NHNDINKKLISQTSR 596
 Qy 630 KDSGVDTCSTTLNEEHSKDPVTPWPSKDGRLIGRLILINKRLKDGSPVPSGAMLG 689
 Db 597 AESPLAASSFLSPDDDTQKN-----RRDGVGVNSLQRTSLDDVAPP----- 643
 Qy 630 LKVGKMTESGRCAFTTKVKGSLADTVGHLRPGDVEWNGRLGQATFEVYNIL 749
 Db 644 -----VAPISKMN-----GH-----IV 655
 Qy 750 ESKPEQV-----ELVSRPGLDIPRIEDTH--AQLESSESPESQKDRPSISVSPMS 804
 Db 656 SSEPSTSTSNQHTSVPIPTVFVVEEKEKAITASTASEPAGVTEPLDDIDENLEP 715
 Qy 805 -----GMLRDV-----POFLSQSLIKLWFKVGHQJLVITLIGARDLPSREDG 847
 Db 716 KHASRRGVRGVRCLLCSFKRSKQSLGSLTILTYHSADKCLAKWHLIRAKNLKAMD 775
 Qy 848 RPRNPKYKIFLDRSDNKRRTKTKLEPKNQTIFYSPVHRFRFRMLEITLWDQ 907
 Db 776 GFSDPYVFLHLLPGNTKATKLTSTIKTILNPENNEMSYGITEDDKKKILRVTLDR 835
 Qy 908 ARVREESFEGLILETALLDDEPHYKLTQHDVSLPLPR 951
 Db 836 DRI-----GSDFLGETIALK-KLNDNEMKKFNLYL--SALPVPQ 873

RESULT 6
 TRHY_HUMAN
 ID TRHY_HUMAN STANDARD; PRT; 1898 AA.
 AC Q07283;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Trichohyalin.
 GN THH OR TRHY OR THL.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93280194; PubMed=7685034;
 RA Lee S.-C., Kim I.-G., Marekov L.N., O'Keefe E.J., Parry D.A.D.,
 Steinert P.M.;
 RT "The structure of human trichohyalin. Potential multiple roles as a
 functional EF-hand-like calcium-binding protein, a cornified cell
 envelope precursor, and an intermediate filament-associated (cross-
 linking) protein."
 RT J. Biol. Chem. 268:12164-12176 (1993).
 RN [2]
 RP SEQUENCE OF 1731-1898 FROM N.A., AND CHARACTERIZATION.
 RX MEDLINE=93315897; PubMed=7686953;
 RA O'Keefe E.J., Hamilton E.H., Lee S.-C., Steinert P.M.;
 RT "Trichohyalin: a structural protein of hair, tongue, nail, and
 epidermis."

J. Invest. Dermatol. 101:658-719 (1993).
 - FUNCTION: INTERMEDIATE FILAMENT-ASSOCIATED PROTEIN THAT ASSOCIATES
 IN REGULAR ARRAYS WITH KERATIN INTERMEDIATE FILAMENTS (KIF) OF THE
 INNER ROOT SHEATH CELLS OF THE HAIR FOLLICLE AND THE GRANULAR
 LAYER OF THE EPIDERMIS. IT LATER BECOMES CROSS-LINKED TO KIF BY
 ISODIPEPTIDE BONDS. IT MAY SERVE AS SCAFFOLD PROTEIN, TOGETHER
 WITH INVOLUCRIN, IN THE ORGANIZATION OF THE CELL ENVELOPE OR EVEN
 ANCHOR THE CELL ENVELOPE TO THE KIF NETWORK. IT MAY BE INVOLVED IN
 ITS OWN CALCIUM-DEPENDENT POSTSYNTHETIC PROCESSING DURING TERMINAL
 DIFFERENTIATION.
 - SUBUNIT: Monomer (Probable).
 - TISSUE SPECIFICITY: FOUND IN THE HARD KERATINIZING TISSUES SUCH AS
 THE INNER ROOT SHEATH (IRS) OF HAIR FOLLICLES AND MEDULLA, AND IN
 THE FILIFORM PAPILLAE OF DORSAL TONGUE EPITHELIUM (PROBABLE).
 - DEVELOPMENTAL STAGE: EXPRESSED DURING LATE DIFFERENTIATION OF
 THE EPIDERMIS.
 - DOMAIN: CONSISTS OF NINE DOMAINS. DOMAIN 1 CONTAINS TWO EF-HAND
 CALCIUM-BINDING DOMAINS. DOMAINS 2-4, 6, AND 8 ARE ALMOST
 ENTIRELY ALPHA-HELICAL, CONFIGURED AS A SERIES OF PEPTIDE-STRANDED
 OF VARYING REGULARITY, AND ARE THOUGHT TO FORM A SINGLE-STRANDED
 ALPHA-HELICAL ROD STABILIZED BY IONIC INTERACTIONS. DOMAIN 6 IS
 THE MOST REGULAR AND MAY BIND KIF DIRECTLY BY IONIC INTERACTIONS.
 DOMAINS 5 AND 7 ARE LESS WELL ORGANIZED AND MAY INDUCE FOLDS IN
 THE MOLECULE. DOMAIN 9 CONTAINS THE C-TERMINUS, CONSERVED AMONG
 DIFFERENT SPECIES.
 - PTM: SUBSTRATE OF TRANSGLUTAMINASE. SOME 200 ARGININES ARE
 PROBABLY CONVERTED TO CITRULLINES BY PEPTIDYLARGININE DEIMIDASE.
 - SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE S-100
 FAMILY.
 - SIMILARITY: CONTAINS 2 EF-hand calcium-binding domains.
 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 EMBL; L09190; AAA65582.1; --
 PIR; A45973; A45973.
 HSSP; P02633; A1CB.
 Genew; HGNC:11791; THH.
 MIN; 190370; --
 GO; GO:0005856; C:cytoskeleton; NAS.
 GO; GO:0005509; F:calcium ion binding activity; TAS.
 InterPro; IPR001751; CABP_S100.
 InterPro; IPR002048; EF-hand.
 Pfam; PF00036; ehand; 1.
 Pfam; PF01023; S_100; 1.
 ProDom; PD003407; CABP_S100; 1.
 PROSITE; PS00018; EF_HAND; 1.
 PROSITE; PS00303; S100_CABP; 1.
 Kcatinization; Calcium-binding; Repeat; Citrullination.
 DOMAIN 1 91 S-100 LIKE
 CA_BIND 22 33 EF-HAND 1 (LOW AFFINITY) (POTENTIAL).
 CA_BIND 62 73 EF-HAND 2 (HIGH AFFINITY) (POTENTIAL).
 DOMAIN 314 390 6 X 13 AA TANDEM REPEATS OF
 REPEAT 314 326 R-R-E-Q-E-E-R-R-E-Q-Q-L.
 REPEAT 327 339 1-1 (APPROXIMATE).
 REPEAT 340 351 1-2 (APPROXIMATE).
 REPEAT 352 364 1-3 (APPROXIMATE).
 REPEAT 365 377 1-4.
 REPEAT 378 390 1-5.
 DOMAIN 391 444 9 X 6 AA TANDEM REPEATS OF R-R-E-Q-Q-L.
 REPEAT 391 396 2-1.
 REPEAT 397 402 2-2.
 REPEAT 403 408 2-3.
 REPEAT 409 414 2-4.
 REPEAT 415 420 2-5.
 REPEAT 421 426 2-6.

70 -EMQERIGRLVDRLEMRKRVAGDVNRCLLCEGLGMLGSAC--VVCSCKKKNVCTKG 120

161 GRVSLRNKYVMVNCNLCRMOEBILTKSGAWFYNSGNTLLQPDDQKVPRGLRNEEAPQBKK 220
: : : : :
128 VETNNLRHSV-WLCKICIEQREVKSGAWFFKG-----FPQVLFQ-----PMPJK 173

221 AKLHEQPOFGADGLSVPAVEKRAHGLTRQDTIKNGGVKHQTASDWPSDKKSPVS 280
: : : : :
174 KTKPQQP-----VSEPA-----APEQAFEPKHPA- 198

281 RDQRRYEQSEREDYSQVPDGTMTSPS---DYADRESQRE---PQYPERGHLYRD 335

199 -----RAPRGDSRRGPQCKTGDPDASAPGRNYP 231

336 SNRRGH--RHSKEYIVDDDEVRSDSYERQRREEYQARYSRDPMLARYPVAKPYPEBQM 393

232 PVRRASEARWSS-----SSRDESND----- 252

394 RIHAEVSAKRHRHSJVSLANAELEDRIISLEMDRPQRSVSBERRAAENQBSYNM 453
: : : : :
253 -----HSGAGDSRSRPAQL-----RANSVO 274

454 RTREAQCGSSVPQRTSNHSPPTP---RRSPIPLDR-----PWMRAD---SLRQHILD 501

275 ASRPAPSQSVSPAPPQCPQPGTGGERGGPAGRPFDDQKEVAPSDPQTTPPPREERTG 334

502 -----PSSAVKTKRKDMETMLENDLSLDQSSEVPPPPFRPHKSKGGMQVSLSSSE 556

335 GVGGYPVGAREDMSPSGPYQSAQAAPQAAAQPPPPPEEEEEEAN-----SYDSDE 389

557 BELASTPEYT---SCDDVHELSESVBKQ-----DSQ-----KGRKTSEQVYL 597
: : : : :
390 ATTILGALEFSLLYDQDNSSLQCTIISKAGLKPMDSNGLADPYVKHLHLPAGASKNKLRK 449

598 SDSNTRSERQKRMYYGCHSLEELWSEPOIKOSVDTCSTTYLNEESHSHDKHPVTWQ 657
: : : : :
450 TLRNTRPIWNETLVHGIT-DEDM-----ORKTLAISVCD-----EDKPGHNE----- 492

658 PSKDGEDRLIGRIILINKRLKDGSVPRDSGAMGLKVGGMGBTSGRLCAPITVKKKSGLAD 717
: : : : :
493 -----FIGE-----TFPSKLKLFWORNFNIC--LERVIPMKRAG 526

718 TVGHLRPGDEVLEWNGRLLOGATFEVYNIILESKPQVBLVWSRPIODIPRIDPSTHA 777
: : : : :
527 TTGSAR-----GMALYE-----EQVERV-----GDIEE----- 550

778 QLESSSSPESQKMDRPSISVTSFMSPGMLRDVPQZLSQLSIKLMFDKVGQLIVTILG 837
: : : : :
551 -----ROKILVSLMYSTQOGLVIGIR 573

838 AKDLFSREDGRPNRYVKIYFLPDRSDKNKRKTVKCTLEPKMNQTIYISVPWREREFE 897
: : : : :
574 CVHLAAMDANGYSDPPFVKMLAKDMGKAKHTQIKKTLNFPNEEFFYDIKH-SDLAK 632

898 RMLBITLWDQARVRESEFPLGILILELTALLDBEFHWYK 938
: : : : :
633 KSLDISVMDYD--IGKSN DYIGGCQLGI-SAKERLKHNYE 670

RESULT 8

RF3A_MOUSE

ID RP3A_MOUSE STANDARD; PRT; 681 AA.

AC P47708;

DT 01-FEB-1996 (Rel. 33, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Rabphilin-3A.

RPH3A.

OC Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 44-167 IN COMPLEX WITH RAB3A.
 RC TISSUE=Brain;
 RX MEDLINE=99148269; PubMed=10025402;
 RA Ostermeier C., Brunger A.T.;
 RT "Structural basis of Rab effector specificity: crystal structure of
 RT the small G protein Rab3A complexed with the effector domain of
 RL rabphilin-3A";
 RL Cell 96:363-374 (1999).
 CC -!- FUNCTION: PROTEIN TRANSPORT. PROBABLY INVOLVED WITH RAS-RELATED
 CC PROTEIN RAB-3A IN SYNAPTIC VESICLE TRAFFIC AND/OR SYNAPTIC
 CC VESICLE FUSION. COULD PLAY A ROLE IN NEUROTRANSMITTER RELEASE BY
 CC REGULATING MEMBRANE FLOW IN THE NERVE TERMINAL.
 CC -!- SUBUNIT: Monomer.
 CC -!- TISSUE SPECIFICITY: Specifically expressed in brain.
 CC -!- SIMILARITY: Contains 2 C2 domains.
 CC -!- SIMILARITY: Contains 1 FYVE-type zinc finger.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U12571; AAC62662.1; -
 CC FIR; I58166; I58166.
 CC PDB; 12BD; 12-APR-99.
 CC PDB; 3RDB; 23-DEC-99.
 CC InterPro; IPR000008; C2.
 CC InterPro; IPR002149; LRI.
 CC InterPro; IPR003315; RPH3A effector.
 CC InterPro; IPR001565; Synaptotagmin.
 CC InterPro; IPR000306; Znf_FYVE.
 CC Pfam; PF00168; C2; 2.
 CC Pfam; PF02318; RPH3A effector; 1.
 CC PRINTS; PRO0360; C2DOMAIN.
 CC PRINTS; PRO0399; SYNAPTOTAGMIN.
 CC SMART; SM00239; C2; 2.
 CC PROSITE; PS00499; C2 DOMAIN 1; 2.
 CC PROSITE; PS50004; C2 DOMAIN 2; 2.
 CC PROSITE; PS50178; ZF_FYVE; 1.
 KW Repeat; Synapse; Protein transport; Zinc-finger; 3D-structure.
 FT DOMAIN 1 277
 FT RAB P25/SMG P25A BINDING (GTP GAMMA-S-
 FT BOUND).
 FT ZN FING 88 145
 FT DOMAIN 280 364
 FT PRO-RICH
 FT DOMAIN 384 488
 FT C2 DOMAIN 1.
 FT DOMAIN 542 645
 FT HELIX 50 84
 FT TURN 85 85
 FT STRAND 93 93
 FT STRAND 100 100
 FT TURN 103 104
 FT STRAND 108 110
 FT TURN 112 114
 FT STRAND 117 119
 FT TURN 120 121
 FT STRAND 123 125
 FT STRAND 135 137
 FT STRAND 138 149
 FT TURN 150 150
 FT HELIX 152 155
 FT TURN 156 156
 SQ SEQUENCE 684 AA; 75832 MW; 05838BC3C7A86444 CRC64;
 Query Match 3.2%; Score 262.5; DB 1; Length 684;
 Best Local Similarity 18.8%; Pred. No. 1.4e-05;
 Matches 177; Conservative 105; Mismatches 298; Indels 363; Gaps 38;
 QY 42 AVMDRQKEEKEEQSVLKKEHQKQPTQFFPSGITELVNVNLPQOK-QNNEKEPTK 100
 Db 35 AQTDQRKQEE-----LTDEK-----EINRVIARAERQETMEQERIGR 74

QY 101 LHOQFEMTKQVKKMGHE-----SQOQEQKGDAPTCGLCHKTTPADGCHNCSCYQTRFC 156
 Db 75 LVDRLETKRKNVAGVNRCLCGEQLGMLSA---CVVCE-----CKKQVC 119
 QY 157 ARCGRVSLRSKNVWVNCMLCKQOEILTKGAMFYNSGNTLQOPDQKQVGRLENEBP 216
 Db 120 TKQGVETSNRPHFPVWLCKICLEQREVMKRSQAMFF-----KGPQKQVLP 164
 QY 217 QEKKAKLHQEQFQCAPCDLSVPVAVEKRAHCLTQDTIKNGSGYVGHQIASMDSPDRRS 276
 Db 165 QMPFKTKTPQPAGEPA-----TQEQTPESRH 193
 QY 277 FSVSRDQNRRYEQSEREDYSQYVSDGTMPSPS--DYADRSORR-----PQFYEPGHL 331
 Db 194 PA-----RAPAGDEMRARAPQKPGDLISAPRG 224
 QY 332 NY-----RDSNRGRHRSKYIVDDDEVSERDYEYRQ--RREBYQARY 373
 Db 225 SHGPPTREARSMSTTTTRDSGMDHGHG-----CGAGVTSRSPGGEQCLARRANSVQAS- 278
 QY 374 RSDPNLARYPVKQPB-----YEQMRIHAEVSEARERHSDVSLANA 416
 Db 279 RPAPASMPSPAPPQVQPGPGGSAAPGQRFPEQT-----EAPPSDFGYPCA 328
 QY 417 ELEDRIISLLRMDRPSRQSVSERRAAMENQSYSMERTREAAQOSYSPQTSNHSPTTP 476
 Db 329 V-----APARE-----HTGPTGQFQAAPHTAGPSQAAP 358
 QY 477 RRSPIPLDRPDMRRADSLRKQHLDPSSAVAKTKRKMETMLRNDLSLSD-QSRSVRPPP 535
 Db 359 ARQPPPAEHEE-----EANSYDSQA---TTGALFSLYLDQDNSNLQCTIIRAKG 408
 QY 536 PREHKS KGGKMRQVSLSSSEELASTPEYTS CDDVLESESVSEKSGKGRKTSQG 595
 Db 409 LKPMDSN-----GLADPVVKKLHL-----GASKNKRLTK--- 439
 QY 596 VLSDNTRSERQKRMYYGSHLEEDLSEBFPQINDSGVDTCSSTTLNREHSHSDKHPVT 655
 Db 440 --TLNTRNPNVNETLQYHGIT-EDM-----QRKTLAISVCD-----EDKFGHE- 482
 QY 656 WQPSKQDRLIGRILLNRLKNDGVSFRDSGAMGLKVVGGKMTBSGRICAPITVKYKGS 715
 Db 483 -----FICE-----TFSLKCLKANQKNFNIC--LERVIPMKR 514
 QY 716 ADTCHLRPGDVLWNGRLIQGATFEVYNILLESKEPQVELVWSPIGDIPIRDPST 775
 Db 515 AGTTGSAR-----GMALYE-----EBOVE-----RIGDIEE----- 540
 QY 776 HAQLESSSSSFESQKMDRPSISVTSPMSGMLRDVQPLSQLSIKLWFKVGHQIIVTI 835
 Db 541 -----RGLVLSLWYSTQOGLIIVI 561
 QY 836 LGAKDLPSPDRPRNPVVKLYFLPDRSDKRRKRTTKTKLEPMWNTFYISPVHRRFP 895
 Db 562 IRCVHLAAMDANGYSDPFVVLKLPDMCKAKHKTKQIKKTLNFEENEFTDKH-SDL 620
 QY 896 RERMLEITLWDOARVRESESEFLGHEILLETALLDDEPHVK 938
 Db 621 AKSLDISVNDYD--IGKNDYIGCCQIGI-SAKGERLKHWE 660

RESULT 10

R23A_BOVIN

ID_R23A_BOVIN STANDARD; PRT; 704 AA.

AC Q06846;

DT 01-FEB-1996 (Rel. 33, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Rabphilin-3A.

GN RPH3A.

OS Bos taurus (Bovine).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

RA Kinzler K.W., Nilbert M.C., Su L.-K., Vogelstein B., Bryan T.M.,
 RA Levy D.B., Smith K.J., Preisinger A.C., Hedge P., McKechnie D.,
 RA Finnleir R., Markham A., Groffen J., Boguski M.S., Altschul S.P.,
 RA Horii A., Ando H., Miyoshi Y., Miki Y., Nishisho I., Nakamura Y.,
 RT "Identification of FAP locus genes from chromosome 5q21.",
 RL Science 253:661-665(1991).
 RN [2]
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RC TISSUE=Fetal brain;
 RX MEDLINE=91330307; PubMed=1678319;
 RA Joslyn G., Carlson M., Thilveris A., Albertsen H., Gelbert L.,
 RA Samowitz W., Groden J., Stevens J., Spirio L., Robertson M.,
 RA Sargeant L., Krapcho K., Wolff E., Burt R., Hughes J.P.,
 RA Warrington J., McPherson J.D., Wasmuth J., Le Paslier D.,
 RA Abderrahim H., Cohen D., Leppert M., White R.,
 RT "Identification of deletion mutations and three new genes at the
 RL familial polyposis locus.",
 RL Cell 66:601-613(1991).
 RN [3]
 RP ASSOCIATION WITH CATENINS.
 RX MEDLINE=94082295; PubMed=8259519;
 RA Su L.-K., Vogelstein B., Kinzler K.W.,
 RT "Association of the APC tumor suppressor protein with catenins.",
 RL Science 262:1734-1737(1993).
 RN [4]
 RP DISEASE.
 RX MEDLINE=95174843; PubMed=7661930;
 RA Hamilton S.R., Liu B., Parsons R.E., Papadopoulos N., Jen J.,
 RA Powell S.M., Krush A.J., Berk T., Cohen Z., Tetu B., Burger P.C.,
 RA Wood P.A., Tagi F., Booker S.V., Petersen G.M., Offerhaus G.J.A.,
 RA Termette A.C., Giardiello F.M., Vogelstein B., Kinzler K.W.,
 RT "The molecular basis of Turcot's syndrome.",
 RL New Engl. J. Med. 332:839-847(1995).
 RN [5]
 RP DISEASE.
 RX MEDLINE=97094176; PubMed=8940264;
 RA Eccles D.M., van der Luijt R.B., Breukel C., Bullman H., Bunyan D.,
 RA Fisher A., Barber J., du Boulay C., Primrose J., Burn J., Fodde R.,
 RT "Hereditary desmoid disease due to a frameshift mutation at codon 1924
 RL of the APC gene.",
 RL Am. J. Hum. Genet. 59:1193-1201(1996).
 RN [6]
 RP DISEASE.
 RX MEDLINE=20243021; PubMed=10782927;
 RA Couture J., Mitri A., Lagace R., Smits R., Berk T., Bouchard H.-L.,
 RA Fodde R., Alman B., Bapat B.,
 RT "A germline mutation at the extreme 3-prime end of the APC gene
 RT results in a severe desmoid phenotype and is associated with
 RT overexpression of beta-catenin in the desmoid tumor.",
 RL Clin. Genet. 57:205-212(2000).
 RN [7]
 RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 2-55.
 RX MEDLINE=20384842; PubMed=10926498;
 RA Day C.L., Alber T.,
 RT "Crystal structure of the amino-terminal coiled-coil domain of the
 RT APC tumor suppressor.",
 RL J. Mol. Biol. 301:147-156(2000).
 RN [8]
 RP X-RAY CRYSTALLOGRAPHY (3.1 ANGSTROMS) OF 1021-1035 IN COMPLEX WITH
 RC CTNNB1.
 RX MEDLINE=21564054; PubMed=11707392;
 RA Eklof Spink K., Fridman S.G., Weis W.I.,
 RT "Molecular mechanisms of beta-catenin recognition by adenomatous
 RT polyposis coli revealed by the structure of an APC-beta-catenin
 RT complex.",
 RL EMBO J. 20:6203-6212(2001).
 RN [9]
 RP X-RAY CRYSTALLOGRAPHY (1.90 ANGSTROMS) OF 2034-2049 IN COMPLEX WITH
 RC AXIN.
 RX MEDLINE=20271867; PubMed=10811618;
 RA Spink K.E., Polakis P., Weis W.I.,
 RT "Structural basis of the axin-adenomatous polyposis coli
 RT interaction.",

RL EMBO J. 19:2270-2279(2000).
 RN [10]
 RP REVIEW ON VARIANTS.
 RX MEDLINE=94154728; PubMed=8111410;
 RA Nagase H., Nakamura Y.,
 RT "Mutations of the APC (adenomatous polyposis coli) gene.",
 RL Hum. Mutat. 2:425-434(1993).
 RN [11]
 RP VARIANTS FAP.
 RX MEDLINE=91335211; PubMed=1651563;
 RA Nishisho I., Nakamura Y., Miyoshi Y., Miki Y., Ando H., Horii A.,
 RA Koyama K., Utsunomiya J., Baba S., Hedge P., Markham A., Krush A.J.,
 RA Petersen G.M., Hamilton S.R., Nilbert M.C., Levy D.B., Bryan T.M.,
 RA Preisinger A.C., Smith K.J., Su L.-K., Kinzler K.W., Vogelstein B.,
 RT "Mutations of chromosome 5q21 genes in FAP and colorectal cancer
 RT patients.",
 RL Science 253:665-669(1991).
 RN [12]
 RP VARIANTS FAP.
 RX MEDLINE=93265030; PubMed=1338904;
 RA Miyoshi Y., Nagase H., Ando H., Ichii S., Nakatsuru S., Aoki T.,
 RA Miki Y., Mori T., Nakamura Y.,
 RT "Somatic mutations of the APC gene in colorectal tumors: mutation
 RT cluster region in the APC gene.",
 RL Hum. Mol. Genet. 1:229-233(1992).
 RN [13]
 RP VARIANTS FAP.
 RX MEDLINE=93244793; PubMed=1338691;
 RA Nakatsuru S., Yanagisawa A., Ichii S., Tahara E., Kato Y.,
 RA Nakamura Y., Horii A.,
 RT "Somatic mutation of the APC gene in gastric cancer: frequent
 RT mutations in very well differentiated adenocarcinoma and signet-ring
 RT cell carcinoma.",
 RL Hum. Mol. Genet. 1:559-563(1992).
 RN [14]
 RP VARIANT FAP TRP-1348, AND VARIANTS ASP-1118; MET-1292; VAL-1304 AND
 RP SER-2502.
 RX MEDLINE=93250848; PubMed=1338764;
 RA Nagase H., Miyoshi Y., Horii A., Aoki T., Petersen G.M.,
 RA Vogelstein B., Maher E., Ogawa M., Maruyama M., Utsunomiya J.,
 RA Baba S., Nakamura Y.,
 RT "Screening for germ-line mutations in familial adenomatous polyposis
 RT patients: 61 new patients and a summary of 150 unrelated patients.",
 RL Hum. Mutat. 1:467-473(1992).
 RN [15]
 RP VARIANT FAP TRP-99.
 RC TISSUE=Peripheral blood lymphocytes;
 RX MEDLINE=95134544; PubMed=783149;
 RA Dobbie Z., Spycher M., Huerliman R., Ammann R., Ammann T., Roth J.,
 RA Mueller A., Mueller H., Scott R.J.,
 RT "Mutational analysis of the first 14 exons of the adenomatous
 RT polyposis coli (APC) gene.",
 RL Eur. J. Cancer 30A:1709-1713(1994).
 RN [16]
 RP VARIANT FAP GLY-722.
 RX MEDLINE=95135430; PubMed=7833931;
 RA Stella A., Montero M., Resta N., Marchese C., Susca F., Gentile M.,
 RA Romio L., Pilia S., Prete F., Marenzi C., Guanti G.,
 RT "Four novel mutations of the APC (adenomatous polyposis coli) gene in
 RT FAP patients.",
 RL Hum. Mol. Genet. 3:1687-1688(1994).
 RN [17]
 RP ERRATUM.
 RA Stella A., Montero M., Resta N., Marchese C., Susca F., Gentile M.,
 RA Romio L., Pilia S., Prete F., Marenzi C., Guanti G.,
 RL Hum. Mol. Genet. 3:1918-1918(1994).
 RN [18]
 RP VARIANT FAP ILE-171.
 RX MEDLINE=97144176; PubMed=8990002;
 RA van der Luijt R.B., Meera Khan P., Vasen H.F.A., Tops C.M.J.,
 RA van Leeuwen-Cornelisse I.S.J., Wijnen J.T., van der Klift H.M.,
 RA Plug R.J., Griffioen G., Fodde R.,
 RT "Molecular analysis of the APC gene in 105 Dutch kindreds with

RT familial adenomatous polyposis: 67 germline mutations identified by
 RT DGB, PTT, and southern analysis.;
 RT Hum. Mutat. 9:7-16(1997).
 RN [19]

RP VARIANTS COLORECTAL CARCINOMA TTR-880; IL8-890 AND VAL-1508.
 RX MEDLINE-98080146; PubMed-9415979;
 RA Miyaki M., Nishio J., Konishi M., Kikuchi-Yanoshita R., Tanaka K.,
 RA Muraoka M., Nagato M., Chong J.-M., Koike M., Terada T., Kawahara Y.,
 RA Fukutome A., Tomiyama J., Chuganji Y., Monoi M., Utsunomiya J.,
 RT "Drastic genetic instability of tumors and normal tissues in Turcot
 RT syndrome.;"
 RL Oncogene. 15:2877-2881(1997).
 RN [20]

RP VARIANTS LYS-1307.
 RX MEDLINE-98400248; PubMed-9731522;
 RA Redston M., Nathanson K.L., Yuan Z.Q., Neuhausen S.L., Satagopan J.,
 RA Wong N., Yang D., Nafa D., Abrahamson J., Ozcelik H.,
 RA Antin-Ozerkis D., Andrulis J., Daly M., Pinsky L., Schrag D.,
 RA Gallinger S., Kaback M., King M.-C., Woodage T., Brody L.C.,
 RA Godwin A., Warner B., Weber B., Foulkes W., Offit K.,
 RT "The APC I1307K allele and breast cancer risk.;"
 RL Nat. Genet. 20:13-14(1998).
 RN [21]

RP VARIANTS LYS-1307 AND GLN-1317.
 RC TISSUE=Peripheral blood.
 RX MEDLINE-98393712; PubMed-9724771;
 RA Frayling I.M., Beck N.E., Ilyas M., Dove-Edwin I., Goodman P.,
 RA Pack K., Bell J.A., Williams C.B., Hodgson S.V., Thomas H.J.M.,
 RA Talbot I.C., Bodmer W.F., Tomlinson I.P.M.,
 RT "The APC variants I1307K and E1317Q are associated with colorectal
 RT tumors, but not always with a family history.;"
 RL Proc. Natl. Acad. Sci. U.S.A. 95:10722-10727(1998).
 RN [22]

RP VARIANTS LYS-1307.
 RX MEDLINE-98400259; PubMed-9731533;
 RA Woodage T., King S.M., Wacholder S., Hartge P., Struwing J.P.,
 RA McAdams M., Laken S.J., Tucker M.A., Brody L.C.,
 RT "The APC I1307K allele and cancer risk in a community-based study of
 RT Ashkenazi Jews.;"
 RL Nat. Genet. 20:62-65(1998).

Query Match 3.1%; Score 258.5; DB 1; Length 2843;
 Best Local Similarity 17.9%; Pred. No. 0.00014;
 Matches 304; Conservative 238; Mismatches 618; Indels 535; Gaps 69;

QY 30 SHLTERRKILAVMDRQKREXSVLKIKEHKAQPTWPFPSGITELVNVLPQOQ 89
 DB 1012 NHMDNDGELDTPINYSKYSDQ---LNSGRQSPQNERWAPKHIE--DEIKQSEQ 1065
 QY 90 KQNEKEPQTKLHQFEMTKQVKGMBESQ---QOBKGDAPTCGICHKTKYFADGGH 146
 DB 1066 ROSRNS---TTPVVTSTDDKHLKFQPHFQOE-----CVSPYRSRGANGSET 1112
 QY 147 NCYSQTKFCARCGRVSLRKNVWVNCRLKQKQELLTKGANFNGSNTLQPDQKY 206
 DB 1113 N-----RVGNHGINQVSSLC-----QBDYEDDPTNYSERYSEERQHEE--- 1155
 QY 207 PGLNNEAPQKAKLHQPOFQAGDLSVPAVEKRAHGLTRQDTKNGSGVKHQA 266
 DB 1156 -----EERPTNYSIKYNEKHVQPIDYSL-----KVA 1184
 QY 267 SMPDRKESVSDQRNRYTEREDYSQVPSDGTMPSPSDYADRSQRPQYFE 326
 DB 1185 TDIPSQKQSPFSKSSSQSSKTE-----HMSSENTSTPSSNAKQNLHPSSAQ 1237
 QY 327 EPHGLNYDSNRGRHNSK-----YIVDD-----EDVE 355
 DB 1238 SRSQGFQKAACTKVSINQETITQTCVEDTPTCFGRCSLSLSAEDEICGNTOTBAD 1297
 QY 356 SRDEYQRREYQARYSDPNLARYPVKQYQVPEQRIHAEVSRARHRRHSDVSLAN 415
 DB 1298 SANTLIQIAIEKEKIGTRSAEPVSEVPAVQHPRTKSSRLQGS--SLSESARHKAVEFSS 1356

QY 416 AELEDSRISILMRDP-----SRQSVSERRAAMNQSVSMERTEAQ----- 459
 DB 1357 GAKSPSKGATPKSPKPHVQETPLMFRCTSV8-----SLDSFERSASSVQSPCSG 1412
 QY 460 -----GQSSYPORTNNHSPPTFRSPPIFLDRPDMRADSLRQHELDSPSSAVKTKREK 513
 DB 1413 MVSGIISPDLDPSQQTMPFSPSKTTPP-----PQTA--QTKRE- 1451
 QY 514 METMLANDSLSDQSVSPRPHKSKKGGKMGQVSLSSSEHEL-----ASTPYTSC 568
 DB 1452 ---VPKNAKPTAKKES-----GPKQAAVNAAGRVQVLDPADTLHFATSTPDGPGC 1502
 QY 569 D-----DVEL-----ESRSVSEKGDQO-KGKRT--SEQ 595
 DB 1503 SSSLSALSDEPFQKQVLRTPVQENDNGNETSEQPKSNENQEKASKTIDSEKD 1562
 QY 596 VLSDSN-----TRSRQKRMYYGHSLEEDLSEHSEPOKDSGVDTCS 639
 DB 1563 LLDDSDDDDIIELECIISAMPTTSRKAQPAQATASKLPPPVAKPSQ-----PVYK 1616
 QY 640 TTLNEHSHSDKHPVTWQPSKQDRL--IGRILLN-----KRLDGVSPDQSGMLGL 690
 DB 1617 LLPSORLQPKH-VSFTPGDDMPVYCVGPTPIFSTATSLDLTIESPPNELAAGEGV 1675
 QY 691 KYVGGMTESGRLCAFIKVKKGLADTVGHLRPGDEVLENNRLLQAGATFEVYNILIE 750
 DB 1676 R--GG--AQSG-----EFERDPTIEG--RSTDEA-----OGKTSVTIPELD 1714
 QY 751 SKPEQVQLV--SRPGD-----IPRIPDSTHAQLESSESSSFBSQKDRPSISVT 799
 DB 1715 DNKAEGDILACINSAMPKGSHPKSHKPPFVKLMQDQV-QASASSAPKNQLDGKKKPT 1773
 QY 800 SPMSQMLRDVPOFLSGQLSILKLMFDKVGHLIVILGAKDLPSREDGRPNPVYKVIPL 859
 DB 1774 SPVK-----IPQNTYETRV-----RKNADSKNLNAERVF 1805
 QY 860 PDRSKNRRTKVKATLEPK-----WNQTFIYSPVH-----RR 893
 DB 1806 SINDSKQNLKNSKDFNDKLPNNEDVRGSPAFDSPHHTPTBGTTCFSRNDLSL 1865
 QY 894 EFERMLBITWQARVRESESEFGLIIELETTALLDEPHYKLGTHVSSLPFR-- 951
 DB 1866 DFDDDDVDLSR-EKAELEKAKENKSEAKVTSHLTENQSANIKQA--IAQPIINQ 1922
 QY 952 SPYLPRLQ-----HGESRTELQ-----SKRLSDSEVSDYCE-- 987
 DB 1923 PAFILOKSTPQSSKDIPLDGAATDEKLQFPAIENTPVCFSHNSLSLSLDIDQENNK 1982
 QY 988 -----DGVGVSDYRHNG-----RDLSSTLSVPROVMSNH 1020
 DB 1983 ENEPKETEPPDSQEPSPKQASGVAPKSFEVETPVCFSRNSLSLSIDSEDLQEC 2042
 QY 1021 SPSSGPHVD--VIGTRTSWSPAPPQPNVQGHGTRATGHYNTISMD--RHRVMD 1076
 DB 2043 ISSAMPKIKKQSRLLKGNKHSKSP-----RNN-GGILGEDLTLDLQRPDSEHGLSPD 2095
 QY 1077 HYSDRDEDCAAD-----RQ-----PYRSRSTE 1101
 DB 2096 SENTDKAIQGANSTVSLHQAACLSRQASDSDSLSLKSGISLGSFHLTPQE 2155
 QY 1102 QRLP-----ERTYRSRSPDTN----- 1122
 DB 2156 EKPTSNKGPILLKPGKSTLETKLIESKGIKGGKVKSLTKGRNSHLSGQKQ 2215
 QY 1123 -LMSMPSLMTGRSAPPSPALSRSHPTGSV-----QTPSPSTPGTGRGRQLPPLP 1174
 DB 2216 PLOANPSSIGRWTMIHIPGVNRSSSTSPVSKGPKLCTPAKSPSEB-----QTATSP 2271
 QY 1175 KGTLSRMDILEENRQMLKNTKQVAGSDPLQDTHSKTRSG---WDPHRGAD----- 1226
 DB 2272 GKAPSVKSELSVARQT-----SQIGSS-----KAPSRSGRSDTSPRPAQPLSR 2319
 QY 1227 TVSTKSSDSDVSDVSAVSRSTSSASRFSSTSTMSVQSERPRGNKLSVFTSMQWQMGVS 1286

Query Match 3.1%; Score 258; DB 1; Length 1338;
 Best Local Similarity 19.4%; Pred. No. 5.5e-05;
 Matches 293; Conservative 194; Mismatches 582; Indels 438; Gaps 66;

QY 2 SAPLGRGRPAP-TPAA--SQPPQPEMPDLHLTEERKI-----ILAVMDQ--KK 49
 DB 32 SGPFGRMRKLPMPEAVGTPDTSRMAELEEVTLGKPLQALRVYDLKALQEGRLAK 91
 QY 50 EKEKESVLKIK-----EHAQPTOWPFGGITELV--NNVLOPOKQWNEKEPQT 99
 DB 92 SQKSAVRLKALGALMLNQLKHSHTPHAAFPQNSQIGEMSQNSFT-----KOYLEKQ-QB 146
 QY 100 KLHQPEYMKVQKMGESQOQOEGKDAPTCGICHKTFKADCGCHGNSYCTTKCARC 159
 DB 147 LLEQLERARENAELERASGSEDE-----MTHPEGVA-----SLPPDFQS-- 189
 QY 160 GGRVLSRKNVMMVCMCRKQOEILT-----KSGAFYNGSNTLQOPDQKVRGLNNEA 215
 DB 190 -----SLNPELELSTHSPKSSSFSEKG-----ESDDEKPKGRSSR 230
 QY 216 PQEKAKLHQPOQAPGDLSPVAVEKGRAGHLTRQDTIKNGSGVYKHQIASMPDRKR 275
 DB 231 VQAKSKLPYSQTAEEEDQETPS-----R 256
 QY 276 SPVSERDQRRYKOSEREDYSQVPSDGTMPSPSYADRSQRE-----POPYEPECH- 330
 DB 257 NLVRADRNKLIIEEIEEIEEIEE-----DDDEEEVEDEAKGREAAPTILKQFEBEGSE 313
 QY 331 -----LNYRDSNRGRHSHKEYITVDDDEVEDRDEYERQREBYQ----- 370
 DB 314 RTRAKPEKVDEKPLNIR-SQEKLEKGRVTRSQEARESHLARQOQEKETQIVSLPQ 372
 QY 371 -----ARYRDP-----NLARYPKVQPYEOMRIHAFV-----SRA 402
 DB 373 EENVKSSQLEKSSQSPPLPELDEKAPVVLQ--EQIVSEETPPPLITKRASSPP 430
 QY 403 RHERRHSDVLANAELEDSISL-----LRMDRPSR-QRSVSERRAAMENQRS 449
 DB 431 THIQOEEMFVEGAPPVLIQSLSPNTDAGAREPLASHPAQLLSLSPLSGTDTTKAE 490
 QY 450 YSMERTRE-----AQGSSYQRTSNHSPPTPR--SPIPL-----DR 485
 DB 491 SPAGVSDSVLPLAQ-KSSLPECTQKGVSEERKSAFLTVVEEPAPAKITPEPMKK 549
 QY 486 PMRRAADSLRKQHLDPSAVKTRKMETMLRNDLSLSDQSESVRPPP-----PRPKS 541
 DB 550 QLEQKEGRASHALFPE---HSGKQASDSSSSSSPSSSPSKSPSPSVASRP-QS 605
 QY 542 KKGGMQRQVLSLSSSEELASTPEYTSRDDVLESESVSEKGDQKQKTSQGYLSDSN 601
 DB 606 SPGSQRDGAQARVIANPHERPMGSRSTSSRSRSPSSASSSSRSKSLSPGVSRDSN 665
 QY 602 TRSERQKRMYYGHSLEEDLEWSEFQIK-----DSGVDTCSTTLNEEHSKHPVTWQ 657
 DB 666 T-SYETTDPSG-----QEAAPSGQQLVLEPKKATFASVGRHLSH----- 711
 QY 658 PSKGDRLIGRILLNKLKDS-----VPRDSGAMGLKLVGVGKMTESGR 702
 DB 712 FEPEQHVIOEL-----QPEQSPKCEAEAEAPPAATQPTSETQI-----SHLESER 761
 QY 703 LCAPITTKVKGSLADTVGHLRPGDEVLR-----WGRLLQATYEVVNIILSKP 753
 DB 762 ---THTVSEKEVMTDNTENRPNPEVPPPLFVADQVSNDRPEGGAEE-----BEKK 812
 QY 754 EPQVELVSVRPIGDIPIPDSTHAQLESSSSFE-----SQKMDRPSIS 797
 DB 813 ESSNPKSPKALISV-----SATKQVQAGNDETEGQPRKRWGASTAATQK--KPSIS 865
 QY 798 VTSMPSPQMLKNDVQFLSGQSLIKLWFDKVGHLIVTLIGADLPFSREDGRPN----- 851
 DB 866 ITTESLKSLLIPDKP-LAQGEA-----VVDLHADDQSISEFERNGDGDT 911
 QY 852 -PYVKI-----YFLPDRSDKNKRTKTVK-----TLEPKWQTF 885

DB 912 DKGLAKICRTVTVVPAEQENQREHEHEKEPEALPAPPQVSVVALPPVHEVKKVT 971
 QY 886 IYSPVHERFERM--LEITLMDQARVRESEFGEI-----LIEL- 925
 DB 972 LGDTLTRRSISQKSGVSIITIDDPVTAQVPSPPRKISIVHINLVNLPFTLGQLKELL 1031
 QY 926 -ETALLDDPHFY-KLQTHDVSSLPRLRPSYLPRLQFHG-----ESPTRLRQSKRIS 979
 DB 1032 CRTGLAVEAFWIDKIKSHCFVYTSVEEA-VATRTALGVKWPQSNPKFL----- 1081
 QY 980 EYSDYDCGCVGVSDYRNGEDLQSSSTLSVPQWSSNHCSPSGSPHRVDVIGTRSMS 1039
 DB 1082 -CADIABODEL---DY-HRGLLVDRPSETKAEQCAPRLHPPPP----- 1122
 QY 1040 PSAPPQNRVQGHGRGTRATGHYNTISRMDRHMVD--DHYSSDRDRDCEAADROPYHRS 1097
 DB 1123 FVQPPPHFAREQORRAVREQM-----AKREREMERERTSREDFWDKVRGPRSR 1177
 QY 1098 RSTEQPILLERTTTRSSSRERPDYTLMSPLMTGRGAPPSPALSRSHPRTGS----- 1151
 DB 1178 RSDRR-RKRAKSKKKSEKES-----KAQEPFPAKLDDLFRKTAAPCIYW 1225
 QY 1152 -----VQTSPTPTGRRGROLPOLPPGTLERSANDIHERNQMLKLYQVAGS 1203
 DB 1226 LPLTESQIVQEAQAKERAKERKEHEESQKEREKAEKERNQLEREKRE----- 1281
 QY 1204 DFLPLSQDHYSKYRSCMDPHRGADTVTSKSSDSVDVAVSRTSSASPSSTSYMSVQSE 1263
 DB 1282 HSRERDRERDRGDRERDRERDRERDRERDRERDRERDRERDRERDRERDRERDR 1331
 QY 1264 RPRGNRK 1270
 DB 1332 RDRGGR 1338

RESULT 13
 TPR_HUMAN
 ID TPR HUMAN STANDARD; PRT; 2349 AA.
 AC P1270;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Nucleoprotein TPR.
 GN TPR.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93064711; PubMed=1437155;
 RA Mitchell P.J., Cooper C.S.;
 RT "The human tpr gene encodes a protein of 2094 amino acids that has
 extensive coiled-coil regions and an acidic C-terminal domain.";
 RL Oncogene 7:2329-2333(1992).
 RN [2]
 RP REVISIONS AND CHARACTERIZATION
 RX MEDLINE=95096166; PubMed=7798308;
 RA Byrd D.A., Sweet D.J., Pante N., Konstantinov K.N., Guan T.,
 RA Saphire A.C.S., Mitchell P.J., Cooper C.S., Aebi U., Gerace L.;
 RT "tpr, a large coiled coil protein whose amino terminus is involved in
 activation of oncogenic kinases, is localized to the cytoplasmic
 surface of the nuclear pore complex.";
 RL J. Cell Biol. 127:1515-1526(1994).
 RN [3]
 RP SEQUENCE OF 1-142 FROM N.A.
 RX MEDLINE=88262257; PubMed=3387099;
 RA King H.W.S., Tempest P.R., Merrifield K.R., Rance A.J.;
 RT "tpr homologues activate met and raf.";
 RL Oncogene 2:617-619(1988).
 CC -!- FUNCTION: COMPONENT OF THE CYTOPLASMIC FIBRILS OF THE NUCLEAR PORE
 CC COMPLEX IMPLICATED IN NUCLEAR PROTEIN IMPORT. ITS AMINO TERMINUS

CC IS INVOLVED IN ACTIVATION OF ONCOGENIC KINASES.
CC -I- SUBCELLULAR LOCATION: CYTOPLASMIC SURFACE OF THE NUCLEAR PORE
CC COMPLEX. THE ASSEMBLY OF THE NPC IS A STEPWISE PROCESS IN WHICH
CC TRP-CONTAINING PERIPHERAL STRUCTURES ASSEMBLE AFTER OTHER
CC COMPONENTS, INCLUDING P62.
CC -I- TISSUE SPECIFICITY: HIGHEST IN TESTIS, LUNG, THYMUS, SPLEEN AND
CC BRAIN, LOWER LEVELS IN HEART, LIVER, AND KIDNEY.
CC -I- DISEASE: INVOLVED IN TUMORIGENIC REARRANGEMENTS WITH THE MET, TRK
CC OR RAF GENES.
CC -I- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.;
CC WWW="http://www.infobiogen.fr/services/chronocancer/Genes/TPRID282.html".

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

CC EMBL; X66397; CAA47021.1; --
CC EMBL; Y00672; CAA68681.1; --
CC MIM; 189940; --
CC GO; GO:0005737; C:cytoplasm; TAS.
CC GO; GO:0005643; C:nuclear pore; TAS.
CC GO; GO:0006606; P:protein-nucleus import; TAS.
CC GO; GO:0006606; P:protein-nucleus import; TAS.
CC Coiled coil; Proto-oncogene; Chromosomal translocation;
CC Nuclear protein; Transport.
CC FT DOMAIN 78 360 COILED COIL (POTENTIAL).
CC FT DOMAIN 422 571 COILED COIL (POTENTIAL).
CC FT DOMAIN 575 628 COILED COIL (POTENTIAL).
CC FT DOMAIN 758 805 COILED COIL (POTENTIAL).
CC FT DOMAIN 834 869 COILED COIL (POTENTIAL).
CC FT DOMAIN 934 979 COILED COIL (POTENTIAL).
CC FT DOMAIN 1004 1064 COILED COIL (POTENTIAL).
CC FT DOMAIN 1138 1166 COILED COIL (POTENTIAL).
CC FT DOMAIN 1196 1241 COILED COIL (POTENTIAL).
CC FT DOMAIN 1262 1304 COILED COIL (POTENTIAL).
CC FT DOMAIN 1354 1434 COILED COIL (POTENTIAL).
CC FT DOMAIN 1476 1595 COILED COIL (POTENTIAL).
CC FT DOMAIN 527 530 POLY-SER.
CC FT DOMAIN 1833 1836 POLY-GLU.
CC FT DOMAIN 1957 1964 POLY-ASP.
CC FT DOMAIN 2295 2298 POLY-SER.
CC SEQUENCE 2349 AA; 265600 MW; AFDD6885CEDCA9EF CRC64;

Query Match 3.1%; Score 255.5; DB 1; Length 2349;
Best Local Similarity 18.1%; Pred. No. 0.00015;
Matches 341; Conservative 268; Mismatches 675; Indels 599; Gaps 76;
QY 10 RPAPTAAASQPPPEMPDLSHL-TEERKIIILAVMDRQKEEKEQSVLKIKEHKAQP 68
DB 644 RPSSTQVSTVPVPTVSTEAIRAKALKQIQIFENYKKEAENE----- 690
QY 69 TQWFFSGITELVNNVLPQKQKPEKQ-TKLHQ-----FMYKQVKKQGE-- 118
DB 691 -----KIQNEQLEKQEQVTLRSQNTKLTQDFASKRYEMLQDNVGEVREIT 740
QY 119 SQQQQKQKQDAPTCGICHTKTFADCGCHNCVCTKFCARCGVSLRSKVMVNCILR 178
DB 741 SLHNRQKLTAFT-----Q 754
QY 179 KQBIILTKSGANFNNGSNTLQPDQKVP-RGLRNEEAPQK-----KAKLHPQPOFG 231
DB 755 KQEQII-----NTWQDLRGANEKLAVALAEVNAELKKEMLKSEVPLSQRE--- 803
QY 232 APGLDLPVAVKGRAGHLTRDTI-----KNGSGVGHQIADMDPSDRGRSFSVSRDQNR 286
DB 804 -----SLIAEQGQMLLTNLQTLQGLERSETETKQRLSSQI---EKLHEIHLAKKL 855
QY 287 YEQSEERS-----DYSQVPSDGTMPSPSDYADRDSOREQFTEEPGLNYRD 335

DB 856 ENEVEQRHTLTNLDVQLDTRKOLDTETNLHLNKLKNAQKEIATLQHLSNVEQV 915
QY 336 SNRGHREHSEKIVVDEVEDSESDYERQREHYQARY-----SDNLARYFV----- 384
DB 916 ASQSQRTGKQPSNKEDVDLVSQLRQTEBQVNDLKERLKTSTSNVQYQAWTSLRES 975
QY 385 --KPOVTEQRIHAEVS-----RAREHREHSDVSLANARLEDSR-----LSL 425
DB 976 LNKKEQVTEVRLNLEVLKSAEAPQTLKKEKMEVEKEKQLQDKKRAIESMEQQLSE 1035
QY 426 LMDRPPSQRSVRRRAAMENQRSYMERTRAQSSYPTORTSNHSPPTPRRSPIDLR 485
DB 1036 LKTLTSSVQNEVQALQASTALSNEQARRDCQQAIAVEAQN-----KYERELMLHA 1090
QY 486 PMRADSRLRQ-----HHLDPSSAVRYT-----KRRM-----E 515
DB 1091 ADVEALQAAKQSVQMAASVROHLETTQKAESEKLECKASWEERERMLKDEKVCVCKE 1150
QY 516 TMLRNDLSLSSQSSVRRPPPHKSKKGGKMRQVSLSSSE----- 557
DB 1151 DLEKQRLHDLQLEK-----SDKVVASVTEGVQGFANVLSLSEKSGSQILILPFRREK 1207
QY 558 -----ELASTPEYTSDDVLESESVSEKGDQNGKR-----KTS 592
DB 1208 EIAETFEVAVQVSLRYRQVLELLESLAEKKEVQVTAQMAQHELMKTE 1267
QY 593 EGVGLSDNTSRERQKMYGCHSELEDEJSEFQIKDSGVDTCSSTLAEHSHS--- 649
DB 1268 TMMVMTNOMLREKER-----LEQDLQOMQAKVRLKELDILPQENAEKSEKSGM 1320
QY 650 -----DKHPVTWQPSKQDRLIGRILLKRLKLDKSGVPRDSCGMLGLK 691
DB 1321 LQAEKCLLEEDVKERKARNGHLVSQO--KDPDTEYRKLSEK-----EVHTK 1366
QY 692 VVGKMTESGLCAPITKVKG-----SLADTVGHLAPGDVLEWN----- 732
DB 1367 RIQQLTEEIGELKAEIARSNASLTNNQNLQSLAKEDLMKVKTEKTIQKDLAKIIDIQE 1426
QY 733 -----GRLLQ-----GATFEVYNIILSKPEQVELVVRPIGDIPRIDPST 775
DB 1427 KVTITQVKIKGRYKQYHELKAAQDKVMTSAQSGDHQBOHVSQVEMQEL---KET 1482
QY 776 HAQLESSESSFSQO-----KMDPSISVTSMPSPGMLRVDVQFLSQ-- 817
DB 1483 LNAQETKSKLSQVENLQKLTSEKTEARNLQBTQVQLSELS--RLQDLQRTQOE 1540
QY 818 -----LSKLWFDKVKHQLITVILGAKDLPSREDG--RPRNPYKLYFLPD 861
DB 1541 QLROQITEKEETKRAIVAASKIAH-----LAGVKQDLKENEKELKQNG-----ALDQ 1590
QY 862 RSDKDKERTKVKKTLKPKNQPIYSPVHR--REFPERMLEITLMDQARYREHSEFLG 919
DB 1591 QKELDVRIYALKSQYEGR-----ISRLERLEHREHREHLE-----QDEPQPSFN 1636
QY 920 EI-----LIBLET-----ALLDDEPHWYKQLTHDVSLPLRPSYPLPRQLHGES 965
DB 1637 KVPESQOQITLTKTPASGERGIASDPP-----TANIKPTVPSVPSKYTAAMAGNK 1690
QY 966 PTERLQSKKISSEVSDYDCEGCVGVSDYRNGEDLOSSTLSVPEQVMSNHCSPSGS 1025
DB 1691 STPRASIRPMVTATVT-----NPTTPTATVMTPTQVESQBAWQSEGP 1734
QY 1026 PHRDVILGR-----RSMSPSAPP-----PQNVQEGHGTATGCHYITISEMDRHRVDDH 1077
DB 1735 VEHVVFPGSTSGSVRSFNVQPSISQPILTVOO-----QTOAT----- 1773
QY 1078 YSSDRDRDCEAARQPHRSRSTEQRPFLERT-----TTRSSSRSPDTNMRSM 1127
DB 1774 -----AFVQPTQOS-----HPQIEPANQELSSNIVEVQSSFPVERPETS----- 1812
QY 1128 PSLMTGRSAPPS---PALGRSHPRTSVQVTSSTPTGRR--GRQLPOLPPKGLTERSA 1182
DB 1813 TAVFGTVSATPSSSLPKRTREBEDSTIEASQVSDDTVEMPLPKLKSVPVGT-EEV 1871

1183 MDIERNRKMLNKYQVAGSDPRLQDYHVKYRSG-----NDPRKADTVSTKSDSDSV 1237
 1872 MAEESTDGEVETQVY-----NQDSQDSIGEGVTQDVTMEDSEETS-QSLQIDL 1920
 1238 SDVSAVSRVSSASRFSSTYSMSVQSRPRGRNKRISVFSKMNQKQ-----MGVS 1286
 1921 GFLQSDQQTTS-----QDQGGKGDVVIDSDDEDEDDDDDDDDDDTGMDGE 1970
 1287 GRKLYKTSISGDMCSLEKNDGSDTAVALGTS-GKRRSSIGAKMVAIVGLSRKRS 1345
 1971 GEDSNEGTV-----SADGNDGYEADDARGGDTDPGTETESMG-----GGGNHRA 2017
 1346 ASQLSQTEGGKRLST-----VORSTETGLAVEMNMTRQASRSTGSMNYSSEGNL 1401
 2018 ADSQSGEGNTGAESSEFSQVSRQCPSSASE-----RQAPR-----APQSPRRPFP 2066
 1402 IPPGVRLASDSQFSDFLDGLG-----AQLVGRQTLAPMGCDLQVGMKKGQLE 1452
 2067 LPPRLTIHAPPO-----ELGPPVQRIQNTRRQSVGRGLQLTGGIGQMCHFFDDEDR-- 2118
 1453 VEIIRARGLVNPKSGKTLFAPYVVKVYLLDNGVCIARKKTKVA---RKTLEPLYQQLLSFE 1509
 2119 -----TVPESTPLVVEH-----RTDGPBAIHSPQVAGVFRFRGPP-----E 2156
 1510 ESPQQRVLQIIVGDIYGRMDHKSFMVQAQ-----ILLDELELNMVIGWFKL---FPSSLV 1563
 2157 DMPQTSSSH-----SDLGQLASQGLGMYETPLFLAHEESGCRSVPTTPLQVADVTVT 2212
 1564 DPTSAPLTRASQS-----SLESSSTG 1584
 2213 ESTTSDASEHASQSVFMTVSTG 2235

RESULT 14

APC_RAT
 ID APC_RAT STANDARD; PRT; 2842 AA.
 AC P70478;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Adenomatous polyposis coli protein (APC protein).
 GN APC.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Fischer 344/N; TISSUE=Brain;
 RX MEDLINE=96116966; PubMed=8563176;
 RA Toyota M., Ushijima T., Kakiuchi H., Watanabe M., Imai K., Yachi A.,
 RA Sugimura T., Nagao M.;
 RT "cDNA cloning of the rat APC gene and assignment to chromosome 18.";
 RL Mamm. Genome 6:746-748(1995).
 RN [2]
 RP MUTAGENESIS.
 RC STRAIN=Sprague-Dawley, and Fischer 344/N;
 RX MEDLINE=95148647; PubMed=7846077;
 RA Kakiuchi H., Watanabe M., Ushijima T., Toyota M., Imai K.,
 RA Weisburger J.H., Sugimura T., Nagao M.;
 RT "Specific 5'-GGGA-3'-25'-GGA-3' mutation of the APC gene in rat colon
 tumors induced by 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine.";
 RL Proc. Natl. Acad. Sci. U.S.A. 92:910-914(1995).
 CC -!- FUNCTION: Tumor suppressor. Promotes rapid degradation of CTNNB1
 and participates in Wnt signaling. APC activity is correlated with
 its phosphorylation state (By similarity).
 CC -!- SUBUNIT: Forms homooligomers. Associates with catenins. Binds
 axin (By similarity).
 CC -!- PTM: PHOSPHORYLATED BY GSK-3B (BY SIMILARITY).
 CC -!- SIMILARITY: Contains 7 ARM repeats.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 or send an email to license@isb-sib.ch).

EMBL; D38629; BAA07609.1; -.

HSSP; Q02248; 3BCT.

InterPro; IPR000225; Armadillo.

Pfam; PF00514; Armadillo_seg; 4.

SMART; SM00185; ARM; 5.

PROSITE; PS50176; ARM_REPEAT; 1.

Wnt signaling pathway; Anti-oncogene; Phosphorylation; Coiled coil;
 Repeat.

DOMAIN 1 728 LEU-RICH.

DOMAIN 2 62 COILED COIL (POTENTIAL).

DOMAIN 3 125 COILED COIL (POTENTIAL).

REPEAT 451 493 ARM 1.

REPEAT 503 545 ARM 2.

REPEAT 546 589 ARM 3.

REPEAT 590 636 ARM 4.

REPEAT 637 681 ARM 5.

REPEAT 682 723 ARM 6.

REPEAT 724 765 ARM 7.

DOMAIN 739 2831 SER-RICH.

DOMAIN 1130 1155 ASP/GLU-RICH (ACIDIC).

DOMAIN 1556 1575 ASP/GLU-RICH (ACIDIC).

DOMAIN 1864 1891 HIGHLY CHARGED.

MUTAGEN 523 523 C-8: IN AN IQ-INDUCED COLON TUMOR.

SEQUENCE 2842 AA; 310530 MW; 3CB2E8A34E8F47 CRC64;

Query Match 3.1%; Score 255; DB 1; Length 2842;

Best Local Similarity 19.5%; Pred. No. 0.0002;

Matches 333; Conservative 204; Mismatches 636; Indels 532; Gaps 78;

20 PFPQ-----PEMPLSLHTEERKILAVNDROCKEERQSV-----LK 59

1525 PFPVENDNGNETEPEQESN-----ENQKEVKEPDSKOLLDDSDDDIE 1571

60 IKER--HKAQPTQWPPSGITELVNVLPQPOKPNKEHPQTKLHQPFMYKEQVWNGE 117

1572 ILKECIISAMPTK-----SSKAKKLAQTASKLP---PVPARKPSQLPVY---KLLPS 1618

118 ESQQQQQK-----GDAPTGCGIKKFKPADGCHNCYCT-----KFCARCG 160

1619 QSRLOAKRHSFTPGDDVP-----RVYCVGPTINFSTATSLDLATIESPPNELAAGDG 1672

161 GRVSLRSNKVMVNCNLCKQOEILTKSGAWFNSGNTLQPDQKVPGRGLRNEAPQKK 220

1673 VRASVQSGEF-----EKDITFTEG-----RSTDEARQK 1702

221 AKLHEQPOFGA---PDLSV-----PAVERGRAHGLTR-----QUTKNGSGV-KHQ 264

1703 VSSIAIPDLGSKAEKGDILABCINSALPKRSHKPFVKKINDVQQAQSWTSSTWKNQ 1762

265 IASDMPDRKESPSVSDQRYE-QSEKEDTSQYTPSGTMRSPSDYADRSOREPQ 323

1763 I--DTKKCKFTSPVKMPQNTYTRVRKNTDSKVNVNTEET-----PSDNKDSKQS 1813

324 FYEPGHLNVDNRNRGRHSKEYIVDDDEVDVSRDYEQRREHYQARYSDPLARYP 383

1814 LKNPKDLN-----DKLPDNE-----RVYGGFTFDSPHYAP 1846

384 VKPQYEQMRIRHAFVSRAPHRHSDVSLANAELEDSEIILAMD---RPSQRVSER 440

1847 IEGTPY-----CFSEND-----SLSLDFDDDDVDLSEKAEKAKGKESKDEA 1890

441 RANENQSRYSMEKTRERQACQSSYPTNSHSPPTFRSPFL-----DRPWRADSL 494

1891 KVTCHTEPSSSQSAAKQASTKHP---VNRGPSKPLQSQPTFPQSSKDVDPGATDE 1947

495 RKQHLDPSASVKTREKMETWLRNDSLSS-----DQSESVRPPPPRPHKSK 543

Db 1948 KLONFAIENTPV.....CFSNSLSLSLSDVDQNNNEETGPRVDAEP-ANAQOQ 1997
QY 544 GKKRQVLSSEELASTPEVTSDDVLESESVSEKDGSKGKRTKSEQGLSDSNTR 603
Db 1998 PKKQASGVAKSFVETVPCFSNS-SLSLSLD-----SEDLRECISS 2044
QY 604 SEROKRMYGCHSLEEDLEWSEPQ-----IKSGVDTCSTTLNBEHSHS-DKHPVTM 656
Db 2045 AMPKKRR-----PSRLKGEQWQSPKGVGLAEADLTLDKDIQRPESSEGLSPDSNFDW 2100
QY 657 QPSKQDELIGRILLNKKLKGSDVR-----DSGMLGLKVGKRTESGRICAFITKVKK 712
Db 2101 KAIQEGANSIVSSLHQAAAAACLSRQSSDSLSILK-----SG-----VSL 2144
QY 713 GSLADTVCHLRPGDVLWNG-----RLQCATFEVYNIILSKPEPQVELVWSRPGDI 768
Db 2145 GS-----PHLPDQEKFTSHKPRILLKPE-----KSTLEAK----- 2179
QY 769 PRIPDTHAQLESSESSSPESQK-----MDRPSISVTPSPMGMLRDVPOFLSGQLS 819
Db 2180 -----KIESENKGIKGGKVYKSLITGKIRSNSEISSQKPLQTNMPSISRGTM 2230
QY 820 IKLWFDKVGHLIVTILGAKDLPSREDGRP-RNPVVKIYFLPDRSDKKRRTK-TVKTL 877
Db 2231 I-----HIPGVNSSSTSPVSKGPPKLTASKSPSPSGPVATTSRGTTPAVKSEL 2282
QY 878 EPKNQTIYSPVRRFRFRMLTILWDQARVRESEFLEILLLELTALLDDEPHY 937
Db 2283 SPITQTSIHSGNKGPSRSRST-----PSRPTQO----- 2315
QY 938 KLOTHVSSLPPLRPSPVLPRLQ-----HGSPTRRLQSRKIBDSVSDYDCBGVGV 993
Db 2316 -----PLSRPMQSGNISPGRNGISTPNKLSQLPRTSPSTASTK-SGSGSKM 2364
QY 994 SDYHNGEDLOSSLTSLVPEQVMSNHCSPSPSHRVDVIGTRNSWSPAPPOQVROCH 1053
Db 2365 S-YTSPGRQSQNLK-KOTGLSKN-----ASSIP-----RSESASKGLNKNNSNG- 2409
QY 1054 RGTRATGHYNTISMRDRHVMDHYSDRDRCDAADRPQYHRGRSTQRLLEKTTTRS 1113
Db 2410 -----SNKKVELSRM-----SSTKSSGSE-SDR-----SERPALVQSTFI 2444
QY 1114 RSSRPTDNLMSPLMTGSRAPPS-----PALSRHPRTGVSQVTSSTP 1160
Db 2445 K--EAPSPTLARKLEESASFESLSPSRPDSPTRSAQTPVLSPLDM-SLSTHPSVOA 2501
QY 1161 GTGRGRQLPOLPP--KCTLERS-----ANDIEERNQMKLKYQVAGSDPLE 1208
Db 2502 GWRK-----LPPNLSPTIYSDCRPSKRHDIARSHSESPSRLPVNR-----AGTWKR-- 2549
QY 1209 QDYHSKYRSGMDPHRGADTVSTKSDSDVSVASVRSSTSSASRFSSTSYMSVQSERPRGN 1268
Db 2550 --EHSKH-----SSSLPRVSTWRKTGSSSSILSASSESEKAKSEDE 2589
QY 1269 RKISVFTSKMQRONGVSGKNLTSTSLGDMCKLEKNGDSQSTAVGALTSCKKRSS 1328
Db 2590 KHVNSVPGPROMKENQVPTK-----GTWRKIKESIEISPTNT-VSQTTSSG-----AA 2635
QY 1329 IGAKMVAIVGLSRKSRASQLSQTEGGKKLRSTVQSTETGLAVE-----MENWTRQASR 1385
Db 2636 SGASKTLI-----YQAPAVSRKTEDVWVRIECPINPNSRGSPT 2676
QY 1386 ESTDGSNNYSSEGNLIPPGVRLASDSQFSDFLDGLGPAQVIGRQTTLATPANGDIQVGM 1445
Db 2677 GNTPEVIDSISEKN---PSIKSDKDTQKQSGVSGSPVQTVGLENL----- 2721
QY 1446 DKKQLEVEITRANGLVVKPSKTLPAVYKVYLLDNGVCIATKKTIVARKTLEPLYQOL 1505
Db 2722 --NSFIOVEAPEQGTETKAGQGS-PAPVAET-----GETCWAER-----TFPSSSS 2765
QY 1506 LSPFESQGRVLIQVWGDYGRMDHKSPMGVAQIILDELELSNNVIGFKLPPSSSLVDP 1565

Db 2766 SSKHSSPGTVAARVTPTNTNPSRKSSADSTSA-----RPSQIPTP 2807
QY 1566 TSAPLTRASQS-SLESSTGSPYSR 1589
Db 2808 VGSSTKRDSDTSTESSGAGSPKR 2832
RESULT 15
APC_MOUSE STANDARD; PRT; 2845 AA.
AC Q61315; Q62044;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Adenomatous polyposis coli protein (APC protein) (mAPC).
GN APC.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND VARIANTS.
RC STRAIN=C57BL/6J, and CAST/Ei; TISSUE=Brain;
RX MEDLINE=92263101; PubMed=1350108;
RA Su L.-K., Kinzler K.W., Vogelstein B., Preisinger A.C., Moser A.R.,
RA Luongo C., Gould K.A., Dove W.F.;
RT "Multiple intestinal neoplasia caused by a mutation in the murine
RT homolog of the APC gene";
RL Science 256:668-670 (1992).
[2]
RN ERRATUM.
RP Su L.-K., Kinzler K.W., Vogelstein B., Preisinger A.C., Moser A.R.,
RA Luongo C., Gould K.A., Dove W.F.;
RL Science 256:1114-1114 (1992).
[3]
SEQUENCE OF 1-45 FROM N.A.
RC STRAIN=BALB/c; TISSUE=Liver;
RA Dicker F., Lambertz S., Reitnair A., Ballhausen W.G.;
RT "The murine APC gene: alternative splicing of 5' untranslated
RT region segments";
RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
[4]
ALTERNATIVE SPLICING.
RP MEDLINE=94061824; PubMed=8242607;
RX Oshima M., Sugiyama H., Kitagawa K., Taketo M.;
RA APC gene messenger RNA: novel isoforms that lack exon 7.;
RL Cancer Res. 53:5589-5591 (1993).
CC -!- FUNCTION: Tumor suppressor. Promotes rapid degradation of CTNNB1
CC and participates in Wnt signaling. APC activity is correlated with
CC its phosphorylation state (By similarity).
CC -!- SUBUNIT: Forms homooligomers. Associates with catenins. Binds
CC axin (By similarity).
CC -!- ALTERNATIVE PRODUCTS:
CC Rvntc-Alternative splicing; Named isoforms=4;
CC Name=1;
CC IsoId=Q61315-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q61315-2; Sequence=VSP_004116;
CC Name=3;
CC IsoId=Q61315-3; Sequence=VSP_004117;
CC Name=4;
CC IsoId=Q61315-4; Sequence=VSP_004116, VSP_004117;
CC -!- TISSUE SPECIFICITY: EXPRESSED IN LIVER, SPLEEN, KIDNEY, HEART,
CC LUNG, BRAIN, STOMACH, INTESTINE, TESTIS AND OVARY.
CC -!- PTM: PHOSPHORYLATED BY GSK-3B (BY SIMILARITY).
CC -!- SIMILARITY: Contains 7 ARM repeats.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>)

or send an email to license@sib-sib.ch).

```

CC      EMBL; M88127; AAB59632.1; -
DR      EMBL; U02937; AAA03443.1; -
DR      PIR; I49505; I49505.
DR      HSSP; Q02248; 3BCT.
DR      MGD; MGI:88039; Apc.
DR      GO; GO:0005737; C:cytoplasm; IDA.
DR      GO; GO:0005634; C:nucleus; IDA.
DR      GO; GO:0008013; F:beta-catenin binding activity; IDA.
DR      GO; GO:0016055; P:Wnt receptor signaling pathway; IDA.
DR      InterPro; IPR000225; Armadillo.
DR      Pfam; PF00514; Armadillo_seg; 4.
DR      SMART; SM00185; ARM; 5.
DR      PROSITE; PS00176; ARM_REPEAT; 1.
KW      Wnt signaling pathway; Anti-oncogene; Phosphorylation;
KW      Alternative splicing; Repeat; Coiled coil.
FT      DOMAIN 1 61
FT      COILED COIL (POTENTIAL).
FT      DOMAIN 125 245
FT      COILED COIL (POTENTIAL).
FT      DOMAIN 1 728
FT      LEU-RICH.
FT      REPEAT 451 493
FT      ARM 1.
FT      REPEAT 503 545
FT      ARM 2.
FT      REPEAT 546 589
FT      ARM 3.
FT      REPEAT 590 636
FT      ARM 4.
FT      REPEAT 637 681
FT      ARM 5.
FT      REPEAT 682 723
FT      ARM 6.
FT      REPEAT 724 765
FT      ARM 7.
FT      DOMAIN 739 834
FT      SER-RICH.
FT      DOMAIN 1130 1156
FT      ASP/GLU-RICH (ACIDIC).
FT      DOMAIN 1556 1575
FT      ASP/GLU-RICH (ACIDIC).
FT      DOMAIN 1864 1891
FT      HIGHLY CHARGED.
FT      VARSPLIC 243 276
FT      Missing (in isoform 2 and isoform 4).
FT      /FTID=VSP 004116.
FT      VARSPLIC 310 410
FT      Missing (in isoform 3 and isoform 4).
FT      /FTID=VSP 004117.
FT      T -> A (IN STRAIN CAST/ET).
FT      V -> I (IN STRAIN CAST/ET).
FT      Y -> F (IN STRAIN CAST/ET).
FT      A -> T (IN STRAIN CAST/ET).
FT      A -> S (IN STRAIN CAST/ET).
FT      G -> A (IN STRAIN CAST/ET).
FT      H -> Q (IN STRAIN CAST/ET).
FT      T -> A (IN STRAIN CAST/ET).
FT      T -> S (IN STRAIN CAST/ET).
FT      T -> S (IN STRAIN CAST/ET).
SQ      SEQUENCE 2845 AA; 311086 MW; 145CA73CF570A499 CRC64;

Query Match          3.1%; Score 254.5; DB 1; Length 2845;
Best Local Similarity 18.8%; Pred. No. 0.00021;
Matches 362; Conservative 241; Mismatches 692; Indels 633; Gaps 86;

QY      45 DRKKEEKGQSV---LKKKE--HKAQPTW-----PPSGITELVNVV 84
DB      1147 EEEQHEEEERTNYSIKYNEKHVDQPIDYSLKYATDISSQKPSFSKNSAQST- 1205

QY      85 LQPKQKP---NEKEPQTKHQPEYKQVKQKQESQOQKQKQDAPTC---GICHT 138
DB      1206 -KPEHLSPESENTAVPNSAKRONLRPSSAQKRG---QTKRG--TTCKVPSINOET 1256

QY      139 KFADGCGHCNSYC--QTKFC-ARCGRVSLSRNKVMVNCNCRKQETILTKSGAFYNSG 195
DB      1257 -----IQYCVEDTIFCSRCSLSLSGADDEIGCDQTTQAD-----S 1296

QY      196 SNTLQOPQKQVPRGLRNEHAPQEKAKLHQEPQFGAPGDLVSPAVEKG-----RAH 247
DB      1297 ANTLQTAEVKENDVTRSAEDPATE-----VPAVSQNAKPSRLQAS 1338

QY      248 GLTQDTIKSGVKHQIASHMPSDRKSPSVSRDONRERYQSEERDYQVPSDGTMP 307
DB      1339 GLSSEST-----RHNKAVEFSSGAK-SPSKSGAQT-----P 1368

QY      308 RSPSDYADRRSQREPQFVEEPCHLNLYRDSNRGRHSHKEYIVDDVEDVESRDEYERQREE 367
DB      1369 KSP-----PEHYVQETPLVFSRCT-----SVSSLSDFSSRSIAS 1402

```

```

QY      368 EYQARYRS-----DPMARYPVK---POPYHEQMRIHAIVSRAR---HERR 407
DB      1403 SVQSEPCSGWVGIIISPSDLPSQPTWPPSRSKTPPPPTQVQAKREVPSKVPAEKR 1462

QY      408 HS--DVSLANAELE-----DGRISLLRMDRPSRQSVSR 440
DB      1463 ESGPKQTAVNAQVORVQVLVDVDTLLHPATESTPDGFCSSLSLSALSDPEPQKDVLR 1522

QY      441 RAAMENQSYNERTRAQAGSSVPORTSNHSPPTFRSSPIPLDRPM----- 488
DB      1523 IMPVQVNDNGNETESEQEPESNENQKEVKPDEKDLDDSDDDDIILEECIISAMP 1582

QY      489 ----RRADSLRKQHLDPSSAVRTKREKMTML-RNDSLSSDSQSVRP---PPPR--- 537
DB      1583 TKSSRKAKKLAQTASKLPPVARKPSQLPVYKLLPAQNRLQAQKHVSFTPGDVERVVCV 1642

QY      538 ---PHSKKKGKQKQVSLSSSEELAS-----TPYTTSCDVELSESSVSXKGSQ 585
DB      1643 EGTPINFSTATSLSDLTIESPFNELATGQVGRAGIQSGEFERKRDITPTGERSTD---DAQ 1699

QY      586 KKK-----RKTSEQGVLSQ-----SNTRS----- 604
DB      1700 RGLSSIVTPDLDDNKABEGDILACINSAMPKSKSHKPFVKKIMQVQQAASSTSGAN 1759

QY      605 ----ERQKRMYYGHSLEEDLWSEPOIKDSQVDTCSSTTLNREHSHSD---KHPTVM 656
DB      1760 KNQVDTKKKKTSPVKMPQNTYRTRVK---NTDSKVNVTETFTSDNKKSKPSLQ 1815

QY      657 QPSK-----DGRLLIGRILINK----- 673
DB      1816 TNAKAFNEKLPNNEDVRGTFLDPSPHYTPTEGTYCFERNDSLSSLDLDDDDVLSRE 1875

QY      674 --RLKQGSVPRDQAMGLK--VVGKWTESGRLCAPITKVKKSLADTVGHLRQDEVLE 730
DB      1876 KAEIRKGVESKDSAKVTCREPNSSQAAKSQASIKHPANRQSKPVLQKQPTFPOSS 1935

QY      731 WNGRLQAGATFEVYNIILSKPEPQVELVVS-----PIGDIPRIPTDTHAQLESSSS 785
DB      1936 KQDPDGAATDEKQLNLAIENTP-----VCFERNSSLSLSDID-----QENNNKE 1982

QY      786 FESQKMDRPSISVTSFMSPGM-----LRDVPQFSGQLSLKLPW---DKVGHOLI 832
DB      1983 SEPIKEAPANSQGEPSKPOASGYAPKSFHVEDTFVCFERNSSLSLSDIDSEDDLLQSCI 2042

QY      833 VTILGAKDLPSREDGRPNPVYKLYFLDPSDKNR-----TKTKVKTLEPKWN 882
DB      2043 SSAMPKKKPSR-----LKSEBKQSPKRVGGIILAEIDLTLDKLQRPDSE 2088

QY      883 QTF-----IYSFVHRERFERMLRITLWDQARVREHESEFLGEILIE 924
DB      2089 HAPSGSENPDWKAIQGANSIVSSLHQAAAAACL-----RQASD--SDSILS 2137

QY      925 LETALLDDDEPHWYKLTQHDVSSLP-----PPSPVLPRLQHGSEFTRLQSK 974
DB      2138 LKSGISLSP--PFL--TPDQEKPFSTKNKGPRIILKPKGKSTLRAKI--ESENKIKGSK 2192

QY      975 RI-----SDSEVSDYDCDGVGVSDYRNGRDLOSSTLSVPRQVSNHSCP-- 1022
DB      2193 KYVSLITGKIRSNSEISS--QMKQPLPTNPSISRGTM-----THIFLARNSSSSTSPVS 2247

QY      1023 -SGSPHRVDVIGTRSMSPAPPQPNVVEQCHRTATGHTYNTIRMDRH--RYWMDHYS 1079
DB      2248 KKGFLKTPA-----SKSPSEGP--GATTSPTGCTKAGK-SHLSPTTRQTSIGSNGK 2298

QY      1080 SDRDCEADRDOPYHRSRSTEQRLPRTTTSRSRSPDTMLRBNPMTLTCRSAPPS 1139
DB      2299 SSRSGSRDSTPSRP-----TQOPL-----SRPMQSPORN-----SISPGENGISP 2338

QY      1140 PALSRSHPRGTSVQTSSTPTGTR-----GRQLPQ--LPPKGTLSRSMDEIRNROM 1192
DB      2339 PNKLSQLPRTSSPSTASTSGSGKMSYTPSGRQLSQOQLTKOASLEKNASSIPRSAS 2398

```

Search completed: December 12, 2003, 07:15:04
Job time : 62 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 12, 2003, 06:26:30 ; Search time 49 seconds
(without alignments)
3120.576 Million cell updates/sec

Title: US-09-617-099b-1

Perfect score: 8285

Sequence: 1 MSAPLGRGPRPAPTPAASQP.....TRRASQSLESSTGSPYSRS 1590

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 76:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4489.5	54.2	1553	2 T03301	rab3 effector prote
2	1002	12.1	853	2 T29736	hypothetical prote
3	500.5	6.0	743	2 T00634	hypothetical prote
4	475.5	5.7	1212	2 T00332	hypothetical prote
5	437	5.3	1325	2 T25753	hypothetical prote
6	358	4.3	547	2 T34318	hypothetical prote
7	338	4.1	1021	2 S44644	hypothetical prote
8	307	3.7	1898	1 A45973	trichohyalin - hum
9	278.5	3.4	1386	2 T49316	profilaggrin relat
10	272.5	3.3	5327	2 T13564	microtubule-associ
11	265	3.2	1550	2 T42727	proliferation pote
12	266	3.2	2526	2 T20531	hypothetical prote
13	264	3.2	681	2 JX0338	rabphilin-3A - mou
14	262.5	3.2	684	2 I58166	rabphilin-3A - rat
15	261.5	3.2	704	2 A48097	rabphilin-3A - boy
16	258.5	3.1	1944	2 A59438	KIAA1424 protein [
17	258.5	3.1	2843	1 RBHTAP	adenomatous polyo
18	254.5	3.1	2845	2 T49505	adenomatous polyo
19	253.5	3.1	1280	2 T00365	hypothetical prote
20	252	3.0	1791	2 T02345	hypothetical prote
21	246	3.0	1871	2 D96796	probable heat shoc
22	245	3.0	5170	2 T15348	hypothetical prote
23	243.5	2.9	1507	2 B47328	natural killer cel
24	242.5	2.9	2248	2 A35938	profilaggrin - hum
25	241	2.9	1549	1 A40691	trichohyalin - she
26	240	2.9	3147	2 T18674	hypothetical prote
27	238.5	2.9	1017	2 T15598	hypothetical prote
28	238.5	2.9	2722	2 T20532	hypothetical prote
29	237	2.9	2094	2 S33124	tptr protein - huma

30	236.5	2.9	1830	2 A37981	microtubule-associ
31	236.5	2.9	2738	2 E88320	protein F07A11.6 [
32	236	2.8	1825	2 S13507	microtubule-associ
33	234.5	2.8	1323	2 T00037	hypothetical prote
34	234.5	2.8	2938	2 T30249	cell proliferation
35	234	2.8	2207	2 T42759	Wucl3-3 protein -
36	233.5	2.8	1407	1 S28589	trichohyalin - rab
37	232.5	2.8	1558	2 T29253	hypothetical prote
38	231	2.8	1827	2 T16270	hypothetical prote
39	230.5	2.8	2453	2 S60254	nuclear receptor c
40	228	2.8	1217	2 S52714	sericinB - silkw
41	227.5	2.7	2282	2 T42717	DNA-binding protei
42	227	2.7	1359	2 T34036	hypothetical prote
43	227	2.7	1695	2 T19823	hypothetical prote
44	225.5	2.7	1306	2 T13592	hypothetical prote
45	225.5	2.7	1829	2 T26135	hypothetical prote

ALIGNMENTS

RESULT 1

T03301

rab3 effector protein Rim - rat

C/Species: Rattus norvegicus (Norway rat)

C/Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 08-Oct-1999

C/Accession: T03301

R/Wang, Y.; Okamoto, M.; Schmitz, F.; Hofmann, K.; Sudhof, T.C.

Nature 388, 593-598, 1997

A/Title: Rim is a putative Rab3 effector in regulating synaptic-vesicle fusion.

A/Reference number: Z14897; MUID:97394473; PMID:9252191

A/Accession: T03301

A/Status: preliminary; translated from GB/EMBL/DDBJ

A/Molecule type: mRNA

A/Residues: 1-1553 <MAN>

A/Cross-references: EMBL:AF007836; NID:92317777; PIDN:AB66703.1; PID:92317778

A/Experimental source: tissue-type brain

C/Genetics:

A/Note: RIM

C/Function:

A/Description: Rim protein is proposed as Rab3-dependent regulator of synaptic-vesicle

C/Keywords: GTP binding; zinc finger

Query Match	54.2%	Score	4489.5;	DB 2;	Length	1553;			
Best Local Similarity	57.2%;	Pred. No.	4.2e-214;						
Matches	952;	Conservative	209;	Mismatches	319;	Indels	183;	Gaps	37;
QY	1	MSAPLGRGPRPAPTPAASQPPEMPDL	SHLTEREKIILAVNDKQKEEKEQSVLKI	60					
Db	1	MSAVGPRGPRPT-----VPPMQEL	DLSHLTEREIIIMAVNDKQKEEKEEAKLKC	56					
QY	61	KREHKAQP-----TQWPPSGITELVNNVLQ	POQKQPKERKPT--KLHQQFEWYKE	110					
Db	57	VYEDMAKPAACKTPRNASQPPHPLNI	FCVVCVFRPSRSGGPRDRLWLFQFSYKE	116					
QY	111	QVKMGESQSQOQ--EKGDAFTCGICHTK	KFDAGCHNGSYCTKFCACGGVSLRSNK	169					
Db	117	QVKIGEEARRYGHEKIDAPTCGICHTK	KFDAGCGHLSYCTKFCACGGVSLRSNN	176					
QY	170	-----VWVVCNLCKQOEIITKSGM	FTNSGNTLQPDQKVRGLNE-----EAPQE	218					
Db	177	EDKVMVMVNCNLCKQOEIITKSGM	FTNSGNTLQPDQKVRGLNE-----EAPQE	218					
QY	219	KKAKLHEQPOFGAPGLSPVAVKGRANGL	TRDTINKS-----GVKH	263					
Db	231	KKALQERSRQ-----TPLTAAV-----	SSQDTATFANPLRNKGLPQQALGPSQ	279					
QY	264	QIA-----SDMPSDRKSPSVSRDQNR	RYQSKEREDYSQV--PSDQTPSRPSDVADR	317					
Db	280	QKASRSRSPPREKKAPGLS--EQNGKQ	GQSKRKVPKSVQPGEGI-----ADERK	333					
QY	318	SOREPOFYEPGHLYNRDNRGRHSKEYI	VDDDEVDVESRDEYRORREYQARYSDP	377					

334 ERRETRLEKGRSQDSD---RPEKRDNGRAVED-----QKQKRESEYOTRYSDP 381
378 NLARYPVKPYEOMIHAHVSRARHERHSDVSLANALEDSRISLLEMDRPSRQSV 437
382 NLARYPVKPYEOMIHAHVSRARHERHSDVSLANALEDSRISLLEMDRPSRQSV 427
438 SERRAAMENQSRYSMERTREAOQSSYPTTSNHSPTPRSPPTPLDPPOMRADSRLKQ 497
428 AEATAGKRAPATARSPPSPRARAQAQPTETGPPPPRPAQAPSPPEPRVPEPLKQ 487
498 HHLDPSSAV--RKTREKMETLNDLSLSDQSESVPPPRPHSKKKGKQWQVLSLSS 555
488 GRLDPGSAVLRAKRAKRAKRAKRAKRAKRAKRAKRAKRAKRAKRAKRAKRAKRA 547
556 EEELASTPYTSCDDVLESESVSEKGDQKGRKKTSEQVLSDSNTRSRQKRMYYG 615
548 EEEGVSTPEYTSCEVLESESVSEKGDLD-----YY-- 579
616 HSLEEDLEWSEPOIKDGVDTCSSTTLNHERHSHDKHPVTWQPSKODRLIGRILLKRL 675
580 -----WLDP-----ATWHSRSTSPISSHPTWQPSKEDRLIGRILLKRT 620
676 KDSVPDSDGAMGLKVVGGKNTSGRLCAPITVKKGLADTVHLPDQVLEWGR 735
621 ---TMPKESGALLGLKVVGGKNTDLGRLGAFITVKKGLADTVHLPDQVLEWGR 677
736 LOGATFEVNVNIILESKPEPOVELWVRPIGDIPIRIPDSTHAOLSSSSSFESQMDRPS 795
678 LPGATNEVNVNIILESKPEQVEIIVSRPIGDIPIRIPDSTHAOLSSSSSFESQMDRPS 737
796 ISVTSMPSPGMLRVPQFLSGQSLKLVFKVGHQILVITILGAKDLPREDGPRPNPVK 855
738 ISVTSPTSPGALKDAPQVLPQSLSVKLVKVGQHLVNVQLATLPPRVGDRPNPVK 797
856 IYPLPDSNDKRTKTKVKTLEPKMNOTPYSPVHREEREMLEITLNDQAVREERS 915
798 MYFLPDRSDSKRTKTKVKTLEPKMNOTPYSPVHREEREMLEITLNDQAVREERS 857
916 EFLGELLILETALLDEPHYKLOTHDVSLSPLPRSPYLPRLQHGSPTRLOQRKR 975
858 EFLGELLILETALLDEPHYKLOTHDVSLSPLPRSPYLPRLQHGSPTRLOQRKR 917
976 ISDSEVDYDCEGCVGVVS--DYRHNGRDLQSSTLSVPEQVMNHCSPSGSPRVDVIG 1033
918 ISDSDISDYEDVDGIGVVPVGYRASAKESKATTLVPEQRTTHRSVSHPGDDQ 977
1034 RTSSWSPAPPOBNVQGH--RQTRAGHYNTISM--DRRYMDHYSDDSDRCEA 1088
978 RPSRLPNV-PLQSLDEIHPTRRSRSPTRHDSRSPADHRSRHSVSQYSSEPDSELM 1036
1089 ADROPYHRS-----RSTEQRPILLRTTRGRSSRPPDNL---MRSMPSLMT----- 1132
1037 LPRAKRGSAESLHMTSELQPSLDARSASTNCLPDTSLHSPERHSRKSERCSTQK 1096
1133 GRSPAPPALSRSHPTGVSSTPGTGRGRQOLPOLPPK-GTLERAGMDIERNRQ 1191
1097 SRKGTASDA-DRTHROGSPQSPADTSPGRRGRQOLPQVPRSGSTEQASLVVEETRQ 1155
1192 M--KLAKYQV--AGSPRLQDHYSKYRSGWDPHRGADTVSTSSSDSDVSGAVRST 1247
1156 MKVXVHRFKQTGGSGQELDHEQYKYNTHQYRSCDNASAKSSSDSDVSAISRAS 1215
1248 SASRPSSTSVMSQSPRGRNKL SVFTSKMQRQMGVSGNLTSTKSTISGDMCSLEKND 1307
1216 STSLSTSTFMSRQSERPRG--RLSSPTPMQGRMTSGRAIKSTSVSGEITLERN 1273
1308 GSQSDTAVGLTGSKKRRSSIGAKVAIVGLSRKRSASQLSQTEGQKGLRSTVQST 1367
1274 GSQSDTAVGTGAGCKKRRSSLSAKVAIV--SRRSSTSLQSTESGHKKLXSTQST 1331
1368 ETGLAVREWNMTQASRSTDGSMNSYSSEGNLIPQVRLASQSDFLDGLGPAQLV 1427
1332 ETGMAAEMRK-MVRQPSRSTDGINSYSSEGNLIPQVRLASQSDFLDGLGPAQLV 1390

1428 GRGTATPAMGDIQVGMMDKKQLVEITIRARGLVVKPQSKTLPAFYVYVLLDNGVCL 1487
1391 GRGTATPAMGDIQVGMMDKKQLVEITIRARGLVVKPQSKTLPAFYVYVLLDNGVCL 1450
1488 KKKTKVARKTLPYQOOLLSFRESPOGRVLOIIVMGDIYGRMDHKSFMGVAQILLDELS 1547
1451 KKKTKVARKTLPYQOOLLSFRESPOGRVLOIIVMGDIYGRMDHKSFMGVAQILLDELS 1510
1548 NMVIGMFKLPPSSSLVDPTSAPLTRASOSSLESSSTGSPYSRS 1590
1511 SMVIGMFKLPPSSSLVDPTSAPLTRASOSSLESSSTGSPYSRS 1553

RESULT 2
T29736
hypothetical protein T10A3.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C/Accession: T29736
R:Sanson, J., Nhan, M.
submitted to the EMBL Data Library, November 1995
A:Description: The sequence of C. elegans cosmid T10A3.
A:Reference number: Z20675
A/Accession: T29736
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A/Residues: 1-853 <RAN>
A/Cross-references: EMBL:U01035; PIDN:AAB37028.1; GSPDB:GN00028; CESP:T10A3.1
A:Experimental source: strain Bristol N2; clone T10A3
C/Genetics:
A:Gene: CESP:T10A3.1
A:Map position: X
A/Introns: 41/3; 110/1; 151/2; 194/1; 232/1; 342/3; 390/3; 528/1; 551/2; 628/1; 674/2;

Query Match 12.1%; Score 1002; DB 2; Length 853;
Best Local Similarity 29.7%; Pred. No. 4.8e-42;
Matches 300; Conservative 142; Mismatches 252; Indels 316; Gaps 36;

652 HPVTWQPSKDDRLIGRILLKRLKDGSPVDRSGAMGLKVVCKKTESGELCAPITVK 711
42 HPVTWQPSKDDRLIGRILLKRLKDGSPVDRSGAMGLKVVCKKTESGELCAPITVK 96
712 KGSLATVGHLPDQVLEWGRKLLQATFERYNIILESKPEQVLEVVSR-----PIGD 767
97 PGSAVDITGELRPGDVEWNGQSLQNAVYQVDSIAASRYDTSVELIVSRALPGD 156
768 IPRIPDSTHAOLSSS-----SFESQ---KMDRPSI-----SVTSMPSPGML--- 807
157 --DFLMLTSPQSSSAYSPVSPAYSPQFORLPNPDLPDLPALQQLSLFHSQSAVPPH 214
808 -----RDVPEP-----LSGQLSLKLVFKVGHQILV 833
215 NWTLSNRSTSTSYSDVDPDLGVPSNREMOESQAPGTGHIPIGRIIVSVFISHDRQLSV 274
834 TILGAKDLSPREDGPRPNPVKIIYFLPDRSDKNRKTKVAKTLBFAWNTFYISVPHR 893
275 ALVRGFDLPSPDTPNPVYKIFLPDRSEKSRQSAVIAETLAPVNDVFFVYNGLTPE 334
894 EFRMLEITLMDQAVRERSEFLGRIILETALLDDEFFHYKLOTHDVSLS--PLPR 951
335 MLLQVLELTVMYDQF--GTNSFLGELTLDLASVFLDGE-----LSLMLCILVMD 384
952 PSYLPRLQHGSR-----PTRELQSKRISDSKVSVDYDCEGQVSDYHNGEDLQSS 1008
385 DNPLRLKRLKRSYNAPTRPQ-----SELNYD----- 414
1009 SVPEQVMSSNHCSPSGPHRVDVIGTRSWSPSPAPPQVNEQCHRGTRATGTYNTISM 1068
415 -----HSSNY----- 420
1069 DRHRVMDHYSDRDRDCEADRPYHRSRSTEQRPILLRTTTRSRSSRPPDNLMSMP 1128

Db 421 -----DHISON-----IDKQPHHHLAPND-----EENDEYIDDDLENDI 456
 QY 1129 SIWTSRAPPSPALSR-----SHPTGSVQTSPPSTGTGR-----GR-----1167
 Db 457 DLATGGGARKSRTYRREKMGHGCHGYADWTQNHQR-----QSGYTSDRGYGRNMIGRAYN 512
 QY 1168 QLPOLPPKGLTSLRSAMDIEERNRQMKLKYQVAGSDPRLQDYHYSKYRQWDP-----HRG 1224
 Db 513 RQQRPRRSATALSQREDM-----IDPTRKRD 542
 QY 1225 ADTVTSKSDSDVSAVARTSSASRFS--STSYMSVQSERPRGRNKISVFTSKMNRQ 1282
 Db 543 DNEYSMERESTTVN-----RNNNANNTSNDTSF-----AETPTANTVRVPIKETQN-- 591
 QY 1283 MGUSGKLTSTISGMCSLEKNDGSDTAVGALGTSGKKRSSIGAKWAVLGLSRK 1342
 Db 592 -----SLASSSSVAG-----GGSANNVN-----KERRKSLMTRFIPGRGAEGK 630
 QY 1343 SRASQLSQTEGGGKLLRSTVORSTETGL-----AVEMRNWMTQASRESIDGSMNS 1394
 Db 631 -----RTGFARSEEVGIPGNLSSDRLTPTPPFLQASKESTD----- 668
 QY 1395 YSSEGNLIPGVRLASQSDFLDGLGPAQLVGRQTLATPAMGDIQVGMDDKKGQLEVE 1454
 Db 669 -SAHSDNWLP---VLADGPLGTFDNLGPGQVGVGRQVLASPVGLGEIQLMAGRESGIDVE 724
 QY 1455 IIPARGHVLVPGSK-TLPAPYKVLVDNGVCIATKTKTKVAKTLEPLVQOQLSFFESPQ 1513
 Db 725 IIRAKNLVVRPGVKLTSPAPYKVLVMEGRQCIATKANTNAATKTSPLFOQLIENDSPK 784
 QY 1514 GRVLQIVMVDYGRMDHKSFMGVAQILLDELEL-SNNVIGWFKLFPDSSL 1562
 Db 785 KTLQVTLVDYGRMERKVFWSIQIRLEDELSQPLIGYKLFHSSSL 834

RESULT 3

T00634
 hypothetical protein H_DJ0897G10.1 - human
 C:Species: Homo sapiens (man)
 C>Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 05-Nov-1999
 C:Accession: T00634
 R:Kalicki, J.; Elliott, G.
 submitted to the EMBL Data Library, January 1998
 A:Description: The sequence of Homo sapiens PAC clone DJ0897G10.
 A:Reference number: Z14194
 A:Accession: T00634
 A:Status: preliminary;
 A:Molecule type: DNA
 A:Residues: 1-743 <KAL>
 A:Cross-references: EMBL:AC004082; NID:G2822160; PIDN:AAB97937.1; PID:G2822161
 C:Genetics:
 A:Map position: 7q11.23-q21.1
 A:Introns: 36/2; 59/1; 130/3; 172/2; 203/3; 227/3; 288/1; 353/1; 414/1; 461/2; 486/1; 59
 A:Note: H_DJ0897G10.1
 C:Superfamily: protein kinase C C2 region homology
 F:119-233/Domain: protein kinase C C2 region homology <KC2>

Query Match 6.0%; Score 500.5; DB 2; Length 743;
 Best Local Similarity 23.9%; Pred. No. 2.5e-17;
 Matches 221; Conservative 126; Mismatches 306; Indels 273; Gaps 40;
 QY 727 EVLEWGRLLQCATFEVYVNIILSKPEQVELVVRPIGDIPIPIPTSTIAQLSSSSSF 786
 Db 2 QVLEWNGIPLTSKTYEEVQSII--SQSGEABICVRLDNLNLSDSNSOHLHLHPKAV 59
 QY 787 ESQK-----MDRPSISVTSFMSPGML 807
 Db 60 DKAKSPGVDFQLAAELQKLSQSPVLVSVEKSHVHGPTSGAGSSSVSPSPGGS- 118
 QY 808 RDVPQFLSGQLSKLWFDKVGHOLIVTILGAKDLPSSREDGRPNRPYVYKIFLDPDRDNK 867
 Db 119 PSVSKKHGSGSKLQINDY-LGN-LIHLQARNLVPRDNGYSDPFVKVILLPGRGAYK 176

QY 868 RETTKVKTLEPQWQCTFYSPVHRREPRERMLELITLNDQAVRRESESEFGEILIEL-E 926
 Db 177 RTKTHVQKSLNPNWNTVITYKSIEMBLQKLTLEVTVDYDRF--SSNDIFGLVLDLSS 234
 QY 927 TALLDEPFWYKL--QTHDVSSLPPLRPSYLPFRQLHGES--PTRRLQASRLISDSVSD 983
 Db 235 TSHLDNTFRWYPLKEQTESID-----HAKSHSSQSSQSPKPS----- 272
 QY 984 YDCEDGVGVVDYRHRNDRLOQSTLSVPPQVMSNCSFSGSPHRRVDTVIGRTSNPSAP 1043
 Db 273 -----VIKSRSHGIFPDPSKDMQVP--TIKSHSHPGSSKSSSE--GELASHGSPRS 320
 QY 1044 PPQRNVQEGRGTRACHTYNTISMRDRHVRWDHYSSDRDCDAADRQPTHSRSTQR 1103
 Db 321 QSKTSVTQTH-----LED--AGAAATAAAEAQVQ----- 347
 QY 1104 PLLERTVTSRSESRPUTNLMRSPSL-WTGRSAPPSPALSRSHPTGSVQTSPPSTGT 1162
 Db 348 -----LRIQPTAKSGQS-----NHAKQHHHSIAGVLPQ 378
 QY 1163 GRGRQLPOLPPGTLERSAMDIEERNRQMKLKYQVAGSDPRLQDYHYSKYRSGNDP- 1221
 Db 379 RTQSDNLPP-PANGNQDQS-----QLALAKVMSDGPVKP-----EGAKPP 417
 QY 1222 -HRGAD-TVSTKSDSDVSAVARTSSASRFSSTSYMSVQSERPRGRNKISVFTSKMQ 1279
 Db 418 NHRPASSSVSTGSSGSPGSGYSVDSESSSTAGETNLFPF-----PRIGK----- 463
 QY 1280 NRQMGVSGKMLTKSTISGMCSLEKNDGSDTAVGALGTSGKKRSS-----IGAK 1332
 Db 464 ---MGQNGQEPVK-----QPGVGVLADTRAQREPPQCACTFGAR 503
 QY 1333 MVAIVGLSRKRSASOLSOTEGGKLRSTV-----QRSTETGLAVERNWMTQAS 1384
 Db 504 -----NMKEIQMPRIYPTDQGEKYPGSLFPFSLQCLIKT-----NW----- 545
 QY 1385 RESTDGSMSYSSEGNL-IPPGV---RLASDSQP-----SDFLDGLGPAQLV 1427
 Db 546 -KPDVG-----EAQADLELLENLQTERILPPSYLLQTKALQTYKRSIDRSGNGPTNA 599
 QY 1428 GRQTLATPAMGDIQVGM---MDKKG-QLVEIIRARGLVVPGSKT-LPAPYKVVLLDN 1482
 Db 600 TTCFETISVWGEIKALKEMKTDGEQLVLEIQCNITTKSPDHLPLDLYVKIYVMI 659
 QY 1483 GV--CIAKTKTKVAKTLEPLVQOQLSFFESPGRVLQIIVWVDYGRMDHKSFMGVAQIL 1540
 Db 660 STQKVKIKKTRVCRHDSRSPFETFRSLSPAGHSLQILLPSNGCKFMKCKTLIGRACIW 719
 QY 1541 LDELELSNNAVIGWFKLFPDSSLVDPT 1566
 Db 720 LDKVDLRKRIVNHHKL-----LVSPET 740

RESULT 4

T00332
 hypothetical protein KIAA0559 - human (fragment)
 C:Species: Homo sapiens (man)
 C>Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000
 C:Accession: T00332
 R:Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Komura, N.; Ohara, C
 DNA Res. 5, 31-39, 1998
 A:Title: Prediction of the coding sequences of unidentified human genes. IX. The comp
 A:Reference number: Z14086; MUID:98290545; PMID:9628581
 A:Accession: T00332
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: mRNA
 A:Residues: 1-1212 <NAG>
 A:Cross-references: EMBL:AB011131; NID:G3043641; PIDN:BAA25485.1; PID:G3043642
 A:Experimental source: brain
 C:Genetics:
 A:Note: KIAA0559
 C:Superfamily: protein kinase C C2 region homology
 F:965-1088/Domain: protein kinase C C2 region homology <KC2>

Query Match	5.7%	Score 475.5	DB 2	Length 1212
Best Local Similarity	23.3%	Pred. No. 8e-16		
Matches 213	Conservative 137	Mismatches 297	Indels 267	Gaps 39
QY	316	RRSQREPOFYEEPCHLVYRNSRRGHSKXYIVDDDSVSRDYE	---	RRREHEHYQ- 370
DB	363	RRSQEVTFD	---	LAPLOSSRLH
QY	371	---	---	ARYRSDENLARYPVKPPQVYEQ
DB	412	GTBSLDHLAGLSHYHHADTSYRHPFKSEKYSISRLTEKQAAKQLPAAILYQOKSKHKS	---	SRAREHRR 407
QY	408	HSVDVSLAN-AELEDNRISLLEMDPPSQRS	---	VSERRAMENQORSYEMERT
DB	472	LIDPKMSKFPFIOESRD--LEPDYSYMTSTSSIGISSRARLLODITFLGAKNITDQ	---	529
QY	456	---	---	REA-QGQSSYPORTSNHSPPTPRRSP1-PLDRPDMRRADSL
DB	530	QKPMGSSLGTCLGTGNTIISALQDEADKPYSSGSRSPRSPSVYGLDL	---	582
QY	495	RKOHLLDPSSAVRKTTRKEMETMLRNDLSGDSQSESVPPPPRPHKSKKGKMGVQSLS	---	554
DB	583	KR-----DSSSSSLRKAQAEAL--DVFSFHASSTARKTPTSLP16QSRGRIPIVAQNS	---	635
QY	555	SERELAS	---	TPETYSDDVLESESVEKGDSSQKGRKT
DB	636	EESPLPVQVQPMGARAAGLPIISADTRDQFGSSHSPLFEVQOHRMERSTNGYDRDI	---	695
QY	592	---	---	SEQVLSDSNTSRERQKRMYYGHSLEHDSWSBPQIKDSGVDTCSSTLNEEH
DB	696	AFIMDDPQHAMSDEAY	---	HLREBETDM
QY	647	SHSDXHP	---	VTWQP-SKDGRLI
DB	735	GLDRKLPRLVHSRPLSCHOEQIIQNGKTMHYIFPHARIKITRDSXQHTV	---	SGNGLG 791
QY	690	LKVVGGRWT--ESRGLCAFLTIVKKGSLADTVGHLRPQDVLVWNGRLLOQATEPKYVNI	---	747
DB	792	IRIVGKEIPCHSGEIGAYAKILPGSABQTKLMGQWLEWNGIPLTSKTYEEVQSI	---	851
QY	748	ILBSKPEQVELVVRPIGDIPIRIPDTHAQLESSSSSFESQK	---	790
DB	852	I--SQSGEAEICVRLDLNMLSDSENSQHLELHEPPKAVDKAKSPGVDPKQAAELQKVS	---	909
QY	791	---	---	MDRPSISVTSPMSPGM
DB	910	LOQSPVLVSVVEKSGSHVSGPTSGAGSSVSPSQPGSPSVSKKKGSSKPTDGTGVVSH	---	969
QY	813	FLSQLSIKLWFDKVGHOLIIVTILGAKDLSPREGRPNPVKIYFLPDR	---	S 863
DB	970	PITGEIQLOINYD-LGN-LIIHILQARNLVRDNNNGYSDPFVKYLLPGRQVMVQNAS	---	1027
QY	864	DKNGRFTTKVKTLEPKWNOFTYSPVHRREFRERMLEITLMDQARVREESBFLGHILI	---	923
DB	1028	ARYKRRTKQKSLNPENWNTVIKYSIMEQKKKTLFVTVWDYDRF--SSNDFLGEVLII	---	1085
QY	924	EL-ETALLDDEPHWYKL--CTHDYVSSIPLRPPSLPRLQHGES--PTRLRQSKRLSDS	---	979
DB	1086	DLSTSHLNTPRVPLKQTESID	---	HGKSHSQQSSQSPKPS--
QY	980	EVSDDYCDCEGVGVVSYDRYRNGRDIQSSTLSVPEQVMSNNHCSPSGSPHRYVDVLTGRTSWS	---	1039
DB	1128	---	---	VIKSRSHGIPDPSPKDMQVP--TIEKSHSSPGSSKSSSE--GHLSHG
QY	1040	PSAPPPPOPNVEQGH	---	1053
DB	1172	PSRSQSKTSVTOH	---	1185

RESULT 5
T25753
hypothetical protein F4584.3 - *Caenorhabditis elegans*

C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T25753
R;Wilson, R.
submitted to the EMBL Data Library, September 1996
A;Description: The sequence of C. elegans cosmid F4584.
A;Reference number: Z20082
A;Accession: T25753
A;Status: preliminary;
A;Molecule type: DNA
A;Residues: 1-1325 «WIL»
A;Cross-references: EMBL:U70852; PIDN:AB09134.1; GSPDB:GN00022; CESP:F4584.3
A;Experimental source: strain Bristol N2; clone F4584
C;Genetics:
A;Gene: CESP:F4584.3
A;Map position: 4
A;Introns: 25/3; 859/1; 928/1; 966/1; 1002/2; 1106/2; 1167/1; 1255/1; 1274/2

Query Match 5.3%; Score 437; DB 2; Length 1325;
Best Local Similarity 20.2%; Pred. No. 7.2e-14;
Matches 242; Conservative 151; Mismatches 405; Indels 400; Gaps 44;

QY 7 PRCGPATPAASQ-----PPQPEMPDLSHLTHEERKILLAVMD 45
DB 315 PRYHPTSVASHQSPSTTKHMSKSIACQADLPPVKMISLTQLDWRRLPQLATYVG 374
QY 46 ROKKEEKSQVLKIKHEHKAQPTWF-----PFGGITELVNVLQPOKQPNKEBPO--- 98
DB 375 KRSAETQTE-----LANYEITITVYVGSIPRSRRTASRVRBQQQIYNQMPQWND 429
QY 99 -----TKLHQFQFVKYQVKQKHGESSQOQKQDAPTCGICHTKYPADGCGHNCSCYQTK 154
DB 430 LLSBITK--KYFEDYDRLQRFEGEARHHSRRFD-----FHDDDDQD----- 469
QY 155 FCARCGVSLRNLKMYVNCVLCRCKQELITKSGNFTSGNLTLOQPKVPRGLNEE 214
DB 470 -----DMRKNQVME--ELARRKMCAS----- 490
QY 215 APOKKAKLHEQPOFGAPGLSPVAVEKRAHGLTRQTIKNGSGVKGHQLASDMPDRK 274
DB 491 -----CHLSTTDPIFA----- 503
QY 275 RSPSVSRDQNRYYEQSEERDYQYVPSDQTMF-----SPSDYAD--RRSQREPOFYE 326
DB 504 LNPVWMS-----SDYSHVPHYGSLFRIIDYPRGTRSDVRDETTRQQLPQPP 552
QY 327 EPCHLAVDSNRGRHSHKEY--LVDEDVSRDEYERQEEHYQARYR---SDPNLARY 382
DB 553 TWTSYNTGSLPRNFRGFSLDLPPIETIENEGSCFAPPRHRSNLGYESTSMFNLSDPVYLG 612
QY 383 FVPQFQYEEQWRHAHY-----SRARHERHSDV-----SLAN-----A 416
DB 613 DSIPWQHQPVRVYDQIPSGYADTTNLNSLNGVGRSDWVSQYASTLNSQFOSGILQOSA 672
QY 417 ELSDRSILRLMRPGRQSRVSERRAME-----NORSYSMEKTEAQQSSYPORTSN 470
DB 673 QLPQPMIPITRYDALMSDPYVSSRSLVDQMNCYNSQYTHHSTLLAPKSAQ-----WT 727
QY 471 HSPT-----PRSPFILDPRPADSLRQHLDPSSAVRKTREKMTLNDLSLSSD 526
DB 728 HIAPTNTLAQQGIGPONTQM---DPLMMSGRIPSSSSQVYSRNM----- 772
QY 527 QSSVSRPPPPRPHKSKKGKMRQVLSLSSSEELASTPFTYSCDDVLESSSVSEKDSQK 596
DB 773 -----NYGSEPAQSSLE-----EYGNRQYGAAPPTVDVFNSSASD--- 810
QY 587 GKQKTBEGVLSDNTRSERQKRMYYGHSLEEDLEWSEFQIKOSGVDTCS----- 638
DB 811 -WRTTQOGLIQMPQMCCQCHNTE-----DARWPKEDALSRWATATERRAQETA 861
QY 639 -----STTLNEHSHS--DKHPTVWQSPKQDRLIGHILANKRL 675
DB 862 LTSNKNISTGSRYARRPINSRYRNPATNSMPDRH--VARTFAENSRVYDKRILLITRSY 920

Qy 332 NYRDSNRRGHI

D**b** 202 N0TQ0NF7M0N0NS0H0P0N0NHN0M0N0P0N0N0H0N0N0H0G0ANN0H0Q0N0N0RAM0Q0Q0PM 261

```

Db 393 NMERFGRHTAHANRLYSTDDDDSSPSPSTRSTS-----PHSL-----434
QY 392 QMERHAESVRAHERHSDSVLANAELESDRISILAMDRPSRORSVSERRAAMENKSYVS 451
Db 435 -----ATPSSYADTCH-DTSLPADTSDSGVQSDHNPQS-----GLTCS 478
QY 452 MENTRAQGGSSVPTSTNHS-PPTPRSPIDPRDPMRRADSLRKQHHLDPSSAVRNTK 510
Db 479 SSSLTPLQQASH-----DHISGGTTPAR-----ISNPD-----RTTS 511
QY 511 REKMETWLRNDSLSDDQSESVRPPPHKSKGGRQVYSSSSEELASTPYTSRDD 570
Db 512 RV-----AQSAGTSLVTPPPP-----ISSRTSPDNCSSPLAV-----545
QY 571 VELESSESVKSGDSQKGRKT-SEQVLSDSNTRSERQKRMVYGHSLSEEDLEWSEPOI 629
Db 546 --MEHKSSASTASSGGRVVGAEFVLNNHAMNNQ-----NENDINKKLIQTSR 596
QY 630 KDSGVDTCSSTTLNEHSHSDKHPVTWPQSKGDRILGRILLNKRKLDGSPVPSGAMLG 689
Db 597 AESSPLAASSFLSPDDDTQKN-----RRRDGVRVNSQLATSLDDVAPP-----643
QY 690 LKVVGGKMTSGRLCAPITKVKIGSLADTVGHLRPGDEVLEWNGRLLOGATFEVYNIIL 749
Db 644 -----VAPISKNN-----GH-----IV 655
QY 750 ESKPEPOV---ELVVSPIGDIPRIDSTH---AQLESSESSSFOSQMDRPSISVTSFMS 804
Db 656 SSEPTSTTSNQHTSVPIPTVPVPEERBEKALTASTESASPGAVIEPLLDIDENLEP 715
QY 805 -----QMLRDV-----PQFLSGQLSKMLPDKVGHOLIIVILGAKDLPREDG 847
Db 716 KHASRRRGDVGVRGFRCLLCSFRKQSRSLGSLTTLTVHSADKKLAKMLIRAKNKAQMSN 775
QY 848 RPNPVYKIYFLPDRSDKNRRKTKVKTLEPKWNTFIYSPVHRERFRERMLRITLWDQ 907
Db 776 GFSDPYVKFLLPONTKATLTKTIEKTLNPNWEMSYGITEDDKKKILRVTVLDR 835
QY 908 ARVREESREFLGELILETALLADDEPHMYKLOTHDVSSPLPR 951
Db 836 DRI---GSDFLGETRIALK-KLNDNEMKKNVLYL--SALPVQ 873

RESULT 8
A45973
trichohyalin - human
C;Species: Homo sapiens (man)
C;Date: 03-May-1994 #sequence_revision 01-Mar-1996 #text_change 22-Jun-1999
C;Accession: A45973
R;Lee, S.C.; Kim, I.G.; Marekov, L.N.; O'Keefe, E.J.; Parry, D.A.D.; Steinert, P.M.
J. Biol. Chem. 268, 12164-12176, 1993
A;Title: The structure of human trichohyalin. Potential multiple roles as a functional
ed (cross-linking) protein.
A;Reference number: A45973; MUID:93280194; PMID:7685034
A;Accession: A45973
A;Molecule type: DNA
A;Residues: 1-1898 <LEE>
A;Cross-references: GB:L09190; NID:g292835; PIDN:AAA65582.1; PID:g292836
A;Note: authors translated the codon AGG for residue 1714 as Pro
C;Comment: Trichohyalin is a protein of the medulla of the hair and of the inner root sh
Covalent modifications to this protein include conversion of arginine to citrulline and
C;Genetics:
A;Gene: GDB:THH
A;Cross-references: GDB:I36223; OMIM:190370
A;Map position: lq21-lq21
C;Superfamily: trichohyalin; calmodulin repeat homology
C;Keywords: calcium binding; citrulline; EF hand; hair; tandem repeat
F;49-81/Domain: calmodulin repeat homology <EF2>

Query Match 3.78; Score 307; DB 1; Length 1898;
Best Local Similarity 17.48; Pred. No. 3e-07;
Matches 245; Conservative 213; Mismatches 514; Indels 434; Gaps 53;

```

```

QY 19 QPPQPEMPDLHLTEEE--RKIILAVMDRQKEEKEBQSVLAKIEKHIAQPTOMFPFSG 76
Db 264 EFQORQELQE-----EEQLKLELQELRRERQEEQQOQLRRERQOLRRKEE-----312
QY 77 IYELVNVVLPQOKQPNKEPQTKLHQOFYEMKQVYKMGERSQOQO---EQKGDAPTCG 133
Db 313 -----ERREQEERREQERREQERREQERREQERREQERREQERREQERRE-----358
QY 134 ICHKTAFADGCGHNCSCYCTKFCARCGRVSLRSNKMVMVCMCRKQOQILITSGAWFTN 193
Db 359 -----REBQO-----364
QY 194 SGNITLQPDQKVPGLRNEEAPQEKAKLHQPOPGAPGDLSPVAVKGRAGHGLTROD 253
Db 365 -----RREQEERREQERREQERREQERREQERREQERREQERREQERRE-----412
QY 254 TIKNGSVGHQIASDMPDRKESPSVSRDQWRYBQSEER-EDYSQVYPSDGTMPSPSD 312
Db 413 QLRERQERREQERREQERREQERREQERREQERREQERREQERREQERRE-----472
QY 313 YAD-----RREQEERREQERREQERREQERREQERREQERREQERREQERRE-----355
Db 473 RMDLAKREETERHEQERREKQQLARQOHEERERWLKSEERREQERREQERREQERRE-----532
QY 356 SRDE--VERQREVEYQARYTSDPNTLARYPVKQPYEOMRIHAIVSRARHERPHSDVSLA 414
Db 533 RREQLRQERREQLRLRSFQQLAR---EQERLEQLAKREERKRLQERREQL---586
QY 415 NAELEDSRLSLAMDPPSPORSVSERRAAMENORSYSMERTREAOQSSYPTSTNHSPP 474
Db 587 KREQEERDOLLKREERREQERREQERREQERREQERREQERREQERREQERRE-----645
QY 475 TFRSPIDLRDPMRADSLRKQHLDPSAVKTKRKNKMTMLRNDLSDDQSESVRPP 534
Db 646 EREEH-----ELLASEQEERH---EQERREQERREQERREQERRE-----690
QY 535 PRPHKSKGGRQVQVLSSEELA-----STPYTSCDDVLESESVSEKGDQKQK 588
Db 691 LKREHEER---REQELAEQEQAERIKSRIPKW---QWLESEA---DARQSK 737
QY 589 -RTVSEQVLSDSNTSEKQKMYGHSLEDLEWSEPOIKDSGVDTCSSITLNEHS 647
Db 738 VILAPQAGRAEAPQOEKRR---ESELQWE-----EEA 772
QY 648 H-----SDKHPVTWQSKGDRILIGRILNKRKLDGSPVDSGAMLGKLVGKMTESG 701
Db 773 HRCQEEERQRRDFTWQAEKSEKRGQRLSAR-----PPLERQERQERQERQERQER 827
QY 702 RLCAFITKVKKGLADTVGHLRPGDVLWNGRLLOGATFEVYNIILSKPQVYLVV 761
Db 828 FL-----PEEEKEQGR---PEEEKEQGR---QRRERREKELQF 851
QY 762 SRPIGDIPIPDSTHAQLSSSSSFOSQMDRPSISVTSFMSPLMDVPOFLSGOLSLK 821
Db 852 -----LEE-----EQORREAAQOEEEDGLQERREQERREQERRE-----899
QY 822 LMPDKVGHOLIIVILGAKDLPSEDRPNPYKIYFLPDRSDKNRRKTKVKTLEPKW 881
Db 890 KW---RWQL-----EEERKRRHT-LYAKALQERLEKQLQEEHEELQ 931
QY 882 NQTFIYSPVHRERFRERMLRITLMDQARVRESEFLGELILETALLADDEPHMY-----937
Db 932 REE---REKRRQERQYR---EEELQEEHEELRE---ERKKRQERREQYRKDK 981
QY 938 KLOTHDVSSILPPLRPSYPLRQKHESP--TERLQRSKISDSKVSVDCEGCVGVSD 995
Db 982 KLOQKE-----EQLLGEPRRRRQERREKRYTEE-----1011
QY 996 YRHNGRDLQSLTSLVPEQVMSSNHCSFSGSPHVRVDVIGRTGWS---PSAPPPQRNVQ 1052
Db 1012 -----ELQOE-----EQULLREERK-----RRQERQYRKDELQEEBQL 1050

```

microtubule-associated protein homolog - fruit fly (*Drosophila melanogaster*)
N;Alternate names: hypotheical protein EG:4984.1

C;Species: Drosophila melanogaster
 C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
 C;Accession: T13564
 R;Span: L.; Papagiamakis, G.; Siden-Kiamos, I.; Louis, C.
 submitted to the EMBL Data Library, April 1999
 A;Description: Sequencing the distal X chromosome of Drosophila melanogaster.
 A;Reference number: Z17689
 A;Accession: T13564
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-5327 -SPA>
 A;Cross-references: EMBL:AL031128; PIDN:CAA20006.1
 C;Genetics:
 A;Cross-references: FlyBase:FBgn0025392
 A;Introns: 24/2; 52/3; 104/3; 179/1; 232/1; 1669/3; 2566/1; 4798/3; 5272/1
 A;Note: EG:4984.1
 C;Superfamily: Drosophila 576K microtubule-associated protein homolog

Query Match 3.3%; Score 272.5; DB 2; Length 5327;
 Best local similarity 18.5%; Pred. No. 5.7e-05;
 Matches 278; Conservative 216; Mismatches 575; Indels 431; Gaps 58;

QY 24 PEMPDLHLTEERKILAVMDROKKEEKEQSVLKI-----60
 DB 1066 PTLPEDEIRPLDQIKEDLVIEKVKYKEETKEAIVATVQTLPAPLAIDTILASATK 1125
 QY 61 --KEEKAQPTQWPPFSG-----ITELVNNVLPQKQKNEKBPQTKLHQOFM---107
 DB 1126 DAPKDANAELGELPDSGERVLPMTTFEAQNLRLDVINKTPDS-VADLPVHEADLGLY 1184
 QY 108 -----YKEQVKWGESQOQKQK-----GAPT-OGICHTKPKADGCHN 147
 DB 1185 EKDSQDANAKSIGHKESEKEETDDEKENKVKYETELGEPNKKVDISH-----1233
 QY 148 CSYCQTFKFCRCGRVSLR-----SNKVMVNCMLCKQOEILTKSGANFYNGSNTLQ 201
 DB 1234 -----VLLKESVQEAQVVIETVKKQEEIVEA-----TTVITQ 1270
 QY 202 PDQKPRGLNREAPQKAKLHQPOFOGAPGDLSPVAVEKRAHGLTQDTIKNGSV 261
 DB 1271 ENQS-----DLMSQVKDKEHEQ-----KTESGI-----ITEKAQKASATP 1307
 QY 262 KHQIASMPDRKPSVSDQNRRYEQSEREDYSQVPSDGTMPSPS-----DY 313
 DB 1308 EKETSDITSDDELPAQLADPTTVPKSAKDRED-----TGSIESPPTIHEAIEVEQ 1360
 QY 314 ADRESQR-----BQFYEEHGHLYNRSNR--RGHRHSKEVIVDDVDVSKDEYERQRE 366
 DB 1361 AKQEAQKVPAPAEAIKTESPLAKETSPEASATGSVKE---DTEQTKSKSPVSPRPE 1417
 QY 367 EYQARYSDPNLARYPVKQPYVEEQMIRHAEVSRARHRRHSDVSLANAELEDLSILL 426
 DB 1418 SE--AKDKSPFASGEASRPESVAESVKDEAGKAESRRE-----SIATKHDES-----1464
 QY 427 RMDPSPQSVSERRAENQORSVSMETREAOQSSYQORTSNHSPPTPRRSPILDRP 486
 DB 1465 SLKAKQESRRRESLAESIKPESGIDKSLAKESRPSVTDKSKPEPGRRESI-----1519
 QY 487 DMRRADSLRQHLID-----PSSAVR--KTKREKMETLNDLSLSSDQSESVR 532
 DB 1520 -----AESLKAESTKDEKSAAPPKESASRPESVVEVKDETEKSPSRRESI-----AESAK 1571
 QY 533 PP-----PPRPH-----KSKGKGKQVRQVLSSESEELASTPE--YTCDDVLELS--575
 DB 1572 PPIEFREVSRRPESVIDGIDKESAPESRRDPLASKB---ASRPESVLESVKDEPIKSTE 1628
 QY 576 -----ESVSE--KGDQKGRKTSQCVLSDNTSERRQKMYGCHSLEDLEWSPQI 629
 DB 1629 KSRESVAESPKASTKDEKSP-----LTSKDISRPEAVENVM-----DAPFKETSR 1676
 QY 630 KDSGVDTCCSTTLNEEHSKDKHPVTWQPSKDGRLILNKLKAGSVPRDSGAMLG 689
 DB 1677 PESAVGSKDESMSEK-----PSR-----RBSVKDGA-----1703

QY 690 LKAVGKMTESRLCAPIYTKVKGSLADTVGHL-RPGDEVLEWNGRLLOCATFEFYNI 748
 DB 1704 -----AQSETSRSPASVAESANDG--ADDLKELSR-----ESTTQSEAGSI 1744
 QY 749 LSKPDPQVELVVRPIGDIPIRIPDTHAQLSSSSSPESQKMDRPSISVTSFPMSPGMLR 808
 DB 1745 KQEK-SPLASEASRPASVAESVKDEAKSKESRRSVAESPLPSKEASRPASVAE--1801
 QY 809 DVPQFLSGQLSIKMFQKVGHQILVITLGAQD-LPSREDGRPNPNVYKIYFLPDRSDKNK 867
 DB 1802 -----SIKDEAKSKESRRSVAESPLPSKEASRPASVAESIKDEAKSKES 1851
 QY 868 RRTKTVKTLKPKWQTFIYSPVHRE-FRERMLHITLMDQARVREERS--RFLGEILLIE 924
 DB 1852 REESVAEK-----SPLPSKEASRPASVAESIKDEAKSKESRRSVAE--1895
 QY 925 LETALLDDPHWVYKLQTHDVSSLPPLPSPFYLPRLQLHG--ESPTRRLQSKRISDSVYS 982
 DB 1896 -----KSPPLPSKEASRPASVAESIKDEAKSKESRRSVAES 1934
 QY 983 DYDCEDGVGVSDYRHRGRDLQSLSTLSVPEQVMSNHCSPSGSPHRVDYIGTRSWSPA 1042
 DB 1935 PLPSKEA-----SRPASVAESIKDEAKSKES--RRESVAES 1971
 QY 1043 PPOPNVEQCHRGTRATGHTNTIISMDRHRVMDHYSSDRDRDCEADQPTHRSRSTQ 1102
 DB 1972 PLPSK-----BASRPASVAESI-----KDEAKSKESRRSVAEK 2007
 QY 1103 RPLLETTTTRSS-----SERPDTNLRMSPLMTGRSAPPSPALSRSHPTGVTSPS 1157
 DB 2008 SPLPSKEASRPASVAESIKDEAKSKESRRSVAESPLPSKEASRPASVAESIKDEAB 2067
 QY 1158 STPGTCRRGQQLPQLPPKQTLERSANDIERNRQMKLKYQVAGSDPRLQDYHSKYRS 1217
 DB 2068 KSKESRR-----ESAAKSPPLPSKEASRPASVAESVKDEADKSKESRRSMAES 2118
 QY 1218 GMDPHRGATVTKSSDSDVSVASRTSSASRFSSTSYMSVQSRPRGNRKISVFTSK 1277
 DB 2119 G-----XQSIKIGQSPKLEYSRPESVAB-----SVKDDPVK-----SK 2152
 QY 1278 MONRQMGVSGKNLTKSTISGDMCSLEKNDGSDTAVGALGTSGKRRSSIGAKMVAIV 1337
 DB 2153 EFSRRRESVAGS--VTADSRDQSPLESKAGASRPESVDSVKDEAKSKESRRSKESTSVI 2210
 QY 1338 GLSRK-SRSASQLSQTEGQKKLRTVQSRSTETGLAVENRNMTRQASRESTDGSMNYS 1396
 DB 2211 PPKAKDKSPKEVLQPVSMETIREDADQPMKPSQA-----ESRRESIAESIKASS 2261

RESULT 11
 T42727
 proliferation potential-related protein - mouse
 C;Species: Mus musculus (house mouse)
 C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 02-Sep-2000
 C;Accession: T42727
 R;Witte, M.M.; Scott, R.E.
 submitted to the EMBL Data Library, November 1998
 A;Reference number: Z22246
 A;Accession: T42727
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: mRNA
 A;Residues: 1-1560 -WIT>
 A;Cross-references: EMBL:U83913; NID:G3858884; PID:G3858885; PIDN:AAC72432.1
 A;Experimental source: strain Balb/C
 C;Genetics:
 A;Gene: P2P-R
 C;Function:
 A;Description: involved in hnRNP association and Rb1 binding
 C;Superfamily: RING finger homology
 F;57-107/Domain: RING finger homology <RNR>

Query Match 3.2%; Score 266; DB 2; Length 1560;

```

Db 1347 -----RNGKD---SSGKLP-----CTLNPFDLFMKEBLAVGVKE-SAVKPK 1385
Qy 1045 PQNVVQGHGTRATGCHYNTISRMDRHVMDDHYSSDRDRDCBAADQPYRSRST--- 1101
Db 1396 PQ-----LHSSR-----LSSDLTETDEAAFPDYNRSDSSNV 1420
Qy 1102 -----QRFLLERTTTRSS--ERP--DTNLMRSMPSLMTGRSAPPSPAL 1142
Db 1421 SVKEEAVASISKELKNTTEKAKESLTVATASQCADRSQSQSPVSPSRSHSPGSQ 1480
Qy 1143 SRSHPRGVSQTSPPSPGTCRRGRQIPQLPFGKTLERSAMDYERVRQMKLNYKVAG 1202
Db 1481 TRSHSSASSAGSQDSKK-----KKKKKKKKKKKKKKKKKKKK 1522
Qy 1203 SDPRLQDYHSKYRSGMDPHRGADTVTKSSDSVDSVAVS 1244
Db 1523 ADGVEKSKKHKK-----KKKKKKKKKKKKKKKKKKKK 1559

RESULT 12
T20531
hypoetical protein F07A11.6a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C:Accession: T20531; T27776
R:Palmer, S.
submitted to the EMBL Data Library, October 1995
A:Reference number: Z19287
A:Accession: T20531
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2526 <N1>
A:Cross-references: EMBL; Z66511; PIDN: CAB54210.1; GSPDB: GN00020; CESP: F07A11.6a
A:Experimental source: clone F07A11
R:Gajadaty, S.
submitted to the EMBL Data Library, March 1996
A:Reference number: Z20417
A:Accession: T27776
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2526 <N12>
A:Cross-references: EMBL; Z59904; PIDN: CAB54501.1; GSPDB: GN00020; CESP: F07A11.6a
A:Experimental source: clone ZK20
C:Genetics:
A:Gene: CESP:F07A11.6a
A:Map position: 2
A:Introns: 36/2; 92/2; 182/2; 272/3; 344/2; 426/2; 490/1; 541/2; 700/3; 770/3;

Query Match 3.2%; Score 266; DB 2; Length 2526;
Best Local Similarity 17.9%; Pred. No. 4.6e-05;
Matches 298; Conservative 235; Mismatches 592; Indels 540; Gaps 66;

Qy 180 QOEILTSGAWFYNSGNTLQDPQKVPGRGLRHEAPQEKAKLHEQPQCGAPGLSVP 239
Db 908 KHDITISKGP---SNAINALQARSQMSGDPKKSAP-----STFVVRDAGSLVAQ 956
Qy 240 AVEKRAHGLTQDTIKNGSVKHQIASDMP--SDRKSPSV-----SDQN 284
Db 957 IMSNQPLGLKLPRIEKKASALQNIHQPFNSNANGSTTSTSTHQAQFKDKERK 1016
Qy 285 RRYQSEBERDYSQVPSDGTMPSPSDYADRSQREFPQYEEFGLNYRDSNRRGRHS 344
Db 1017 KKEKKEERE-----REARERKKEKTKERKKKKEERAK-- 1053
Qy 345 KEYIVDDVESRDEYRQREHEQYQARYSDPNLARYVPQPYTEQWRIHAE---VSR 401
Db 1054 ----LEDERQERKKEKKERDKKE-----KSKVKKAKKKKKKK 1090
Qy 402 ARHERRS-----DVSLANAEI-----EDSRISLRM-----DPSRQSV 437
Db 1091 KKRKNGSDSDSDSDNDELDDVRSKTKGMTQEEKDQIALLLSKGIFENLKSRRSD 1150
Qy 438 SERRAAMENQRSYMSMT-----REACQSSYPORTSNHSPPTP----- 476

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 12, 2003, 07:16:45 ; Search time 609 Seconds
(without alignments)
485.573 Million cell updates

Title: US-09-617-099B-1

Perfect score:

Sequence: 1 MSAPLGPRGRPAPTPAASQP.....TRRASQSSLESSTGPSYSR 1590

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 684280 secg. 185983659 residues

Total number of hits satisfying chosen parameters:

Minimum DB req length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Local Processing: Minimum Match 0%
Maximum Match 100%

Database :

Published Applications_AA:*

/cgm2_6/ptodata/2/pubpaa/US07_PUBCOMB.psp.*
/cgm2_6/ptodata/2/pubpaa/PCT_NEW_PUB.psp.*
/cgm2_6/ptodata/2/pubpaa/US05_NEW_PUB.psp.*
/cgm2_6/ptodata/2/pubpaa/US06_PUBCOMB.psp.*
/cgm2_6/ptodata/2/pubpaa/US06_PUBCOMB.psp.*
/cgm2_6/ptodata/2/pubpaa/US07_NEW_PUB.psp.*
/cgm2_6/ptodata/2/pubpaa/PCT05_PUBCOMB.psp.*
/cgm2_6/ptodata/2/pubpaa/US08_NEW_PUB.psp.*
/cgm2_6/ptodata/2/pubpaa/US08_PUBCOMB.psp.*
/cgm2_6/ptodata/2/pubpaa/US09_PUBCOMB.psp.*
/cgm2_6/ptodata/2/pubpaa/US09B_PUBCOMB.psp
/cgm2_6/ptodata/2/pubpaa/US09C_PUBCOMB.psp
/cgm2_6/ptodata/2/pubpaa/US09_NEW_PUB.psp.*
/cgm2_6/ptodata/2/pubpaa/US10A_PUBCOMB.psp
/cgm2_6/ptodata/2/pubpaa/US10B_PUBCOMB.psp
/cgm2_6/ptodata/2/pubpaa/US10C_PUBCOMB.psp
/cgm2_6/ptodata/2/pubpaa/US10_NEW_PUB.psp.*
/cgm2_6/ptodata/2/pubpaa/US60_NEW_PUB.psp.*
/cgm2_6/ptodata/2/pubpaa/US60_PUBCOMB.psp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	8285	100.0	1590	15	US-10-180-326-1	Sequence 1, Appl
2	2905	3.5	3051	12	US-10-144-194-62	Sequence 62, Appl
3	279	3.4	3664	15	US-10-177-293-423	Sequence 423, App
4	278.5	3.4	910	15	US-10-177-293-449	Sequence 449, App
5	270.5	3.3	2665	9	US-09-864-761-34248	Sequence 34248, A
6	267.5	3.2	862	12	US-10-339-782-492	Sequence 492, App
7	258.5	3.1	1564	12	US-10-144-198-2	Sequence 2, Appl
8	258.5	3.1	1564	12	US-10-144-198-4	Sequence 4, Appl
9	258.5	3.1	2843	10	US-09-587-482-1	Sequence 1, Appl
10	257.5	3.1	2843	8	US-08-681-219-32	Sequence 32, Appl
11	257.5	3.1	2843	12	US-10-092-138-30	Sequence 30, Appl
12	257.5	3.1	2843	12	US-09-230-111-30	Sequence 30, Appl
13	252	3.0	671	12	US-10-311-626-5	Sequence 5, Appl
14	251	3.0	72	9	US-09-864-761-38199	Sequence 38199, A
15	251	3.0	2440	12	US-10-341-434-236	Sequence 236, App

ALIGNMENTS

RESULT 1

```

US-10-180-326-1
; Sequence 1, Application US/10180326
; Publication NO. US20030049661A1
; GENERAL INFORMATION:
; APPLICANT: Seino, Susumu
; APPLICANT: Shibasaki, Tadao
; APPLICANT: Ozaki, No. US20030049661A1uaki
; TITLE OF INVENTION: Protein Rim2
; FILE REFERENCE: P21573
; CURRENT APPLICATION NUMBER: US/10/180.326
; CURRENT FILING DATE: 2002-06-27
; PRIOR APPLICATION NUMBER: JP 286372/99
; PRIOR FILING DATE: 1999-10-08<160> 5
; NUMBER OF SEQ ID NOS:
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 1
; LENGTH: 1590
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-180-326-1

```

	Query Match	100.0%;	Score 8285;	DB 15;	length 1590;
	Best Local Similarity	100.0%;	Fred. No. 0;		
	Matches 1590;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MSAPLGPRGRPAPTPAASOPPPQPEMPDLSHLTEERKIIILAVMDRQCKEKKESQVLRKI	60		
Db	1	MSAPLGPRGRPAPTPAASOPPPQPEMPDLSHLTEERKIIILAVMDRQCKEKKESQVLRKI	60		
Qy	61	KKEHKQAQPTQWPPFSGITELVNNVLQPOQKQPKESQPTKLHQOFENYKESQVNNKGESESQ	120		
Db	61	KKEHKQAQPTQWPPFSGITELVNNVLQPOQKQPKESQPTKLHQOFENYKESQVNNKGESESQ	120		
Qy	121	QOQEQKGDAPTCGI CHTKYFADGCGNCSYCOTTFPCACCGRRVSLRSNKVMVNCNLCRQK	180		
Db	121	QOQEQKGDAPTCGI CHTKYFADGCGNCSYCOTTFPCACCGRRVSLRSNKVMVNCNLCRQK	180		


```

QY 314 -----ADRSOREPOFY-----EPGLNTRDSNR-----338
Db 401 DGLTDTIVSSVHTSLLSSFEEDTEEVVTSMSBEGEITSDDDEKNKQNKTKTQTSDSSE 460
QY 339 -----RGRHS-----KEYI-----VDDEBVSROBYERORREBYQARYSDPMLARYPVKQPQ 388
Db 461 GKTYSVRHAYVHPYLYSKYYSDDDELTVQRQSTAKBHER-----LLRRQINREK 514
QY 389 YEECMRTHAUSRARHER--RHSDVSLANALEDSDRISILRMDRPSQRQSVSERRAAMEN 446
Db 515 LEEKROKAKTKSSKTKGQGRSSVDLESSTKSLEPKAARIKEVLKERKVKLEKVALSK 574
QY 447 QR-----SYSMERTBAOQO-----SSYPORTSNH-----SPTPT-RRS 479
Db 575 KRKQDSNRVENSXKKQYEEZSDSKETLTKTSEHCEKEKISSKELKHVHAKSEPKAPRL 634
QY 480 PIPDRDMRADSLAKOHLDPSSAV-----RKTR-----EKME 515
Db 635 SESLHVVDENKESKIEREHRRTSTFVIMEGVQOEBTDRDVKRQVERSEICTEHPQOK 694
QY 516 TMLRNDL-LGSDDOSES-----VRPPPRPHSK-----KGGKR-----548
Db 695 STLAKKHLKDDSETPHLSKLLKKEVSKSEKPEREKTPTSEDKLSVKHYKGDCHKYG 754
QY 549 -QVSLSSSEEL-----ASTPYTSCDDVELESSESVEKSDQKQK--RKTSEQV 596
Db 755 DETELSSSEKGLAVERNIQKSOQTKLSSDDKTERKSHRNERKLSVLGDKGKPVSEYII 814
QY 597 LSDNTRSEKQKMYGHSLEBLESWSPOIKDGVDTCSSTTLNEEHSHEKHPVTM 656
Db 815 KTDENVKKNKKE-----RLSAKTKAHSRRS-----SDSKIORSLSGKQHGITL 864
QY 657 Q---PSKDGLR-LIGRILLAKRKGSV-----PRDSGMLGLKVGVGKMTESGRLCAFIT 708
Db 865 QRSSESSEKCDMDSTNDSNLKPEEVVHKEKERTKSLLEKLVLSKSK-----T 916
QY 709 KVKGSLADTVGHLRPGDVLNWRLLQATFEVNIILSKPEPOVELWSR-----763
Db 917 QGKQKVQVET-----ELQEGATKQ-----ATTPKPKDKNTEENDSEK 954
QY 764 -----PIGDIPIRIPDSTHAQLESSESSSPESQWMDRPSISVTPSPGMLRDVPQLSQLS 819
Db 955 QRKSQVEDKPEETGVPVLETSASSASTQKD--SSHRAKLPLAKEKYKSDKOSTSTLE 1013
QY 820 IKLWFDKVGHOLIVITLGAOLPREDGRPNPVYKIVFLPDRSDKKNRKTATVKKTLBP 879
Db 1014 RKL---SDGHSRSLKSSKDIKKDENK-----SDDKQKEVDSSHE--KA 1055
QY 880 KWNQTFIYSPVHRREFRERMLEITLNDQARVRHRESEPLGEILIELESTALLDDEPHWKL 939
Db 1056 RGNSSLMKKLSRLCENRRGSLs---QEMAKGSE-----KL 1089
QY 940 QTHDVSSILPRLPSPLPRLQHGESPRLRQSKR-----ISDSFVSDYDCEDGGVV 993
Db 1090 AANTLST-----PS-----LQRPKSGDMTLIPEQPMEDSPGVENV 1131
QY 994 -----SDYRNG--RDLQSGSTLS-----VPEQVMSNH 1019
Db 1132 FEVSKTQDRNNNSQDIDSENMQTKTSATVQKDELATCTADSTATAPAYKPGGTGVS 1191
QY 1020 CSPSGSPHRVDVIGRTSNWSPAPPQPNVQBGHGRTRANG--HYNTISRM-----DRHR 1072
Db 1192 NSEKHADHRSTLTKMHQISAVSKMNPQEKPIHERGITTEVNIIDSETVHRLMSSAPSENR 1251
QY 1073 VMDHYSDRDRDCEAADRQPHRSRSTEQPILERTTTRSSRSDPTNLWRMPSLWT 1132
Db 1252 VQKRLKNTAAEHVAQGDATLEHSTNLNDSPLSSVTVPLRESYDPPVIFLDFKRTVLE 1311
QY 1133 GRSAAPPSPA-----LSRSHPTG-----SVOTSPSSSTPGTGR-----1165
Db 1312 GSTASTSPADHSALPNQSLTVRESEVLKTSKGGEGFTVDT--PAKASITSKKHIPAH 1370

```

```

QY 1166 -----GROLPQPPKGTTLERSAMDIERN-----ROMKLNKTKQVAGSDPRLBODYHSK 1214
Db 1371 QATLLDGKQKQKVIPLGS--KLATGVIVENENITKEGGLVDMAXCKENDLAEFNLKQTIKAT 1429
QY 1215 YRS-----CWDPHRGADTVST--KSSDSVSDVSA-VSRTSSASRE-----SST 1255
Db 1430 VENKKGGLAGVHVGLNTEKTAETVLKHKRSFGKVDLSDIVERNENSEVDTSGSG 1489
QY 1256 SYMSVQSER-----PRGNRKISVFTS--KMONRQMGVSGKNLTKSTISGD-----1299
Db 1490 SAPSVLHQHQGTEDVATGPRRAEKTSVATSTEGKDKDVTLSPVKAGPATTTSSETROSE 1549
QY 1300 -----MCSLEKNDSQSDTAVGALGTSGKKERSIGAK-----WYAIWGLSR 1341
Db 1550 VALPCTSIADDEGL-----IIGTHSRNNPLHVGAESECTVFAAAEEGAVVTEGFAR 1602
QY 1342 KRSASQLSQTEGGGKKLSTVORSTETGLAV-----EMRNMWTRQASRE 1386
Db 1603 SETFLTSTTEGESGECAVASEDEADL-LAVHAKIENAVNSVVTTEKODDATVSAGSE 1661
QY 1387 STDGSMNSYS-----SEGNLIFPGVRLASDQSFDFLDG-----LGPAG 1425
Db 1662 KCDGSLSRDSEIYVGTITFTISEVESDGVTSAGTEIRAGSISSEEDGSGQNMWMPKK 1721
QY 1426 LVGRQTLAPANG---DIQVGMMDKKGQLEVEIIRAGLVW-----KPGS 1467
Db 1722 ETEGTVTCGABRSDFNVICSVTGAGPRERMVTCAGVGLGNDAPPGT 1771

RESULT 3
US-10-177-293-423
; Sequence 423, Application US/10177293
; Publication No. US20030124128A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Lillie, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Gannavarpu, Manjula
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Mertens, Maureen
; APPLICANT: Myer, Vic
; APPLICANT: Wang, Youzhen
; APPLICANT: Xu, Yongyao
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Monahan, John
; APPLICANT: Meyers, Rachel E.
; APPLICANT: Bast Jr., Robert C.
; APPLICANT: Hortobagyi, Gabriel N.
; APPLICANT: Puztai, Lajos
; APPLICANT: Meric, Funda
; APPLICANT: Sahin, Aysegul
; APPLICANT: Mills, Gordon B.
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-038
; CURRENT APPLICATION NUMBER: US/10/177,293
; CURRENT FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: US 60/299,887
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 60/301,572
; PRIOR FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: US 60/306,501
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: US 60/325,002
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/362,585
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/XXX,XXX
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 506
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 423
; LENGTH: 3664

```

TYPE: PRT
ORGANISM: Homo sapiens
US-10-177-293-423

Query Match 3.4%; Score 279; DB 15; Length 3664;
Best local similarity 20.1%; Pred. No. 1.4e-07;
Matches 370; Conservative 223; Mismatches 676; Indels 570; Gaps 85;

QY 31 HLTEERKIILAVMDROKKEEKEQSVLKIK--BEKKAQPTONFPFSGITELVNNVLPQ 88
DB 371 HGTSEERYGL--VFPQOEOEKALTASKGLFCGMQIEVTANI--GPTESENEFRPL 425
QY 89 OKOPNEKEPOT-----KLHQOF-EMYKOVKMGESQOQKQGD 128
DB 426 DERIDEPHATRTLFTGNLEKTTVTHDLNIPQFGEIVIDIDIKVGVGPQYAFQYICD 485
QY 129 APTGCIHKTKFADG--CGHNCVQCTKFCARCGGVSLSNKMVMVNCNLCRQOEILTK 186
DB 486 --TASVCKATKQMDGEYLGNN-----RLKCGFGKSM-----PT 516
QY 187 SGAWFYNGSGNTLQO-----PDQKVP-----RGLNEEAPQEKAKLHQPOQGA 232
DB 517 NCWMDGLSLSNVSDYLTRHCRCYGVVVKVVFDELKGM-----ALVLYNEIEYAQA 567
QY 233 PGDLSVPAVE--KGRANG-----LTROQTI-----KNGSGVK--HQIASDMSDR 273
DB 568 -----AVKETGKRTGKIGNKIKVDFANRESQLAFYHCEKSGQDIRDFTYEMLAERER 620
QY 274 K-----RSPSVSRDQRRYEQSBEREDYSQVPSDGTMPSPSDYADRRS 318
DB 621 RASVDYNDQTYVESVTPGTPEDSRDYDARGREFYSEWETVQ-----DYTESR- 672
QY 319 QREPOFYEEPHLYNDRNRRGHRHSKEYIVDDDEVSDEYERQRESEYQARYSDPN 378
DB 673 -----YDDP--REYDVR-----NDPEQDIREYSYRQER--- 702
QY 379 LARYFVKPQYEBQMRTHAVSRAR--HERHSDVSLANAELEDRIISLEMDRPSRORSV 437
DB 703 -----ERERERFSDRDHERR-----PIERSQFVHLRPPQSGASPSQ 743
QY 438 SERAAMENORSYMERTRAQGSSYPQRTSNHSPPTPRSPILDRPDMRRADSLRKQ 497
DB 744 AERLPSDSERRLYS-----RSSDRSGSCSLSPRYEKLKRSRLRYTKN 786
QY 498 HHLDPSSAVRKTKEKMETMLRNDLSLSDQSESVRPPPHKSKGQKMRQVSLSS-- 555
DB 789 EKTDKERTFDPVERERERLIRKEVEKTD-----KQKRGKVHSPSSQSET 838
QY 556 --EELASTPY--TSCDDVLESSESVSEKQSGKQ-----KRKTSQGVLSDSNTR 603
DB 839 DQENEREQSPKPRSCN--KLSREKADKEGIAKNRLELMPCVLTVRKEKEGKVID--HTP 895
QY 604 SEROKRMYGHSLEBEDLESEBQIKDSGVDTCCSTTLNEHSHSDKHVPMQPSKQGD 663
DB 896 VEKLKAKL-----DN--DTVKSALDQKQVQSTEPKASDLK-- 931
QY 664 KLIGRIILNKLKDGVSFRDSCAMGLKLVGGKRTESGRLCAFTIKVKGSLADTVGHLR 723
DB 932 -----LESVENK--VPKEKGLSHVEVV-----EKGRU-----KARK 965
QY 724 PGD-----EVLEWNGRLLOQATPE-----EYVNIILSKPBPQV--- 757
DB 966 PEQPADGVSADVLKLEARKERFADSNLKAQKPEVKKSPFEMEDARVLSKQPDVSSR 1025
QY 758 ELVYSRPICDIPRIPDSTHAQLESSSSFESQKMDR-----PSIVTSFMSRP 804
DB 1026 EVILLRE--GEARKP-----VRKEILKRESKIKLDRLNTVASPKDCOELASISVSGSRP 1080
QY 805 GMLRDVPQFLSQLSIKLFWKVGHOLIVTILGAKDLPSDGRDGRPNPVK--IVFL--- 859
DB 1081 S-----SDLQARLG--ELAGSEVENQ-----FVQSKPIFSK-----POLKQLQVLDQ 1123
QY 860 PDRSDKNKRRTKTKVKTLEPKMNQTFIYSP-----VHRREFRERMLEITLMDQAR 909

DB 1124 PEREDVRKNYCSLRDETPTKSGQKSHSVNTEEKIGIDIDHTQSTRKQMBQRRRQOME 1183
QY 910 VREKSESEFLG---EILILELTALL-----DDEPHWYKLTQHDVSSLPFRPS 953
DB 1184 MEIAKSEKFGPKQOVDEYERESLVHEVCKPQDVTDSPPSKKGKMDHVDFTICKRER 1243
QY 954 PTLFRQLH-----GESPTRLQSRKISDSSEVSDYDCEDGVG-----VVSVDYRNGR 1001
DB 1244 NYRSSRQISSEDSERTGSGSVR-----HGSFHEDEDPTGSPRLLSVAGSPKVDK 1293
QY 1002 DLOSSTLSYPEQVMSNHCSPSGSRPHRVVDVIGRTSNSPSAPPPQPNVQOGR--GTRAT 1059
DB 1294 VLPYSNITVRESLKFNPYDSSRRBQAMAKIKLSVLNSEDRLRWDSOMQODAGRFOV 1353
QY 1060 GHYNTTIRMD--RHRVMD-----HYSSDRDRDCEAADRPQYHRSRSTEQRPILLERTTAS 1113
DB 1354 SPNSIILKDSILKRSVRDLFPQEVPSDEDEGE-----HSHSPRASALYB-----SS 1402
QY 1114 RGSERPTNLMRMPSLMTGRSAPPALSRSHPTGSGVOTGSPSTPTGTRGRQOLPQLP 1173
DB 1403 RLS-----FLARDREDKLAERDERLSSSLERN--KFTSPALDKTITPDT----- 1444
QY 1174 PKGTLEP-----SANDIEERNQMKLANKYQVAGSDPR--LEQDVHESKYRSGND 1220
DB 1445 -XALLERAKSLSSREENWSFLWDNSRFANFRNNKDKKVD SAPRPIPSMYHKKKIRTD 1503
QY 1221 PHRGADVTSGSDSDVSDVSAVSTSSASRPSSTSYMSVQSERPRGRNKKISVFTSKMOW 1280
DB 1504 SEQMD-----DKCEDHKEBEQERQELFASRFLLESSIFRODSK-----LQHLERKEED 1552
QY 1281 ROMVSGKNLTSTSTSGDMCS-----LEKND 1307
DB 1553 SDF--ISGRYIGKQTSRGANSITDIOEPVVLPHSRFELTRMQCKEKEKQPKRVEKOE 1611
QY 1308 GQSDTAVGALGTSGK-----KRSSIGAKMVAIVGLSRKSRASQISQTEGGGKLSRV 1363
DB 1612 DTENHPKTPESADENKDSHLKTPPGVPPSVTVTL--ESAPSALEKTTG----- 1659
QY 1364 QRSTETGLAVEMNNMTQASRESTDGSNNYSYS-----SEGNLIPPGV 1406
DB 1660 DKTVEAPLVEETKTEVPATVSEAKPAPAPVPEQLEQVDLPPGADPDKEAMPAGV 1719
QY 1407 RLASDSQSFDFDLGLOPA-----OLVGRQTLATPAM-----GDIQVGMMDX 1447
DB 1720 EGGSGDQPPYLDKAPPTPGASPSQAESNVDPPEPDSQPLSKPAQKSEANEKPAKCPDA 1779
QY 1448 KQOLEVEI---IRAGLIVKPGSKTLPAPVYKVTLLDNGVCIKTKTKVARKTLELYQ 1504
DB 1780 TADAEPDANKAEAPESPQPPASEDL-----VDPFVAAKDKKPKNSKSKTVPQAA 1831
QY 1505 LLSFESPSQGRVLQIIVMGDYGRMDHK-----SFMGVAQILLD--ELELSNMVIGWFKLF 1557
DB 1832 AVSIVEKFTVKSE-----RIDREKLKRSNPSRGEAQKLELKNKEAKIT-----RTA 1879
QY 1558 PPSLVD-----PYSAPL--TRRASQSLESSTGSPYSRS 1590
DB 1880 SKNSAADLHPHPSLPSLSTRERNRVSATMGJHENS 1918

RESULT 4

US-10-177-293-449
Sequence 449, Application US/10177293
Publication No. US20030124128A1
GENERAL INFORMATION:
APPLICANT: Lillie, James
APPLICANT: Glatt, Karen
APPLICANT: Zhao, Xumei
APPLICANT: Gannavarpu, Manjula
APPLICANT: Kamatkar, Shubhangi
APPLICANT: Mertens, Maureen
APPLICANT: Myer, Vic
APPLICANT: Wang, Youzhen


```

RESULT 6
US-10-339-782-492
; Sequence 492, Application US/10339782
; Publication No. US20030165026A1
; GENERAL INFORMATION:
; APPLICANT: Lynx Therapeutics, Inc.
; APPLICANT: Goodman, Laurie J
; APPLICANT: Bowen, Benjamin A
; TITLE OF INVENTION: Identification of Specific Biomarkers for Breast Cancer Cells
; FILE REFERENCE: 37-000110US
; CURRENT APPLICATION NUMBER: US/10/339,782
; CURRENT FILING DATE: 2003-01-08
; NUMBER OF SEQ ID NOS: 495
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 492
; LENGTH: 862
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-339-782-492

Query Match          3.2%; Score 267.5; DB 12; Length 862;
Best Local Similarity 19.7%; Pred. No. 9,9e-08;
Matches 174; Conservative 146; Mismatches 285; Indels 277; Gaps 41;

QY      187 SGAMFYNSGSGNTLAQPDQKVRGLNBEAPKKALEHQEQFQGAPGDLSPVAVERGA 246
Db      2 SGQWFY-----EAKAKRH-----RDKI 18

QY      247 HGLTRQDTIKNGSGVKH-QIASDWPDRKRSPSVSRDQNRREYQ-----SEKRED 295
Db      19 HG-----ADIIIRASMRKRPQIAAEQSKDRENGAKESWVNVNKAFLPPELAGVWEPEED 75

QY      296 YSQVPSDGTW-PRSPSDYADRRSQREPOFYEEPCHLNVRDSNRRGHRHSKEYIVDDDEV 354
Db      76 AAPASPSSVVNPASVIDMGSQZNRKKNVSPFK-RKPPFNSSKULPGEHSSQOTKNBSQK 134

QY      355 ESRDEYERQRREBYQARYRSDPNLARYPVKPYQPYEEQMRTHARYSRARHERHSDVSLA 414
Db      135 NGRGLPQTSKED-----SETLYRNINFPFKSVKIVSGLTIHERISE--KE 266

QY      415 NAELEDGRISLLRMDRPSRQSRVSERRAAM-----ENRQSYSMERT--REAQGS 463
Db      164 -QKLEKSKQTLPLGSLNGSQIKAPIPKARKMIYKSTDLNKDDNQSPPRQRTSLKARGAPR 222

QY      464 YPQRTSNHSPPTPRSPPIPLDRPMDRADSLSRKHLDPSNAVRK--TKREKMETMLRN 520
Db      223 GILKRNSSGSGTD-----SETLYRNINFPFKSVKIVSGLTIHERISE--KE 266

QY      521 DLSLSSDQSESVRPPPPRPHKSKGGMQVSLSSSEBELASTP-----EYTSDDV 571
Db      267 HSLDSSPNSLP-----LKHVRPSAVKDELQSPGLIHGREVGEFVSLESD 314

QY      572 ELESESVSEKQDSQKGRKNTSEQVLSQNSNTRSEQKRMYYGGHSLREDLWSVSPQKD 631
Db      315 RLAKN-GMEDAGDTBEFQ-----SDPKP-SQYRKP-SLEH-----QSTSSPYVSK 355

QY      632 SGVD--TCSSTYLNEEHSHSDKHPVTWP-SKGDRLIGRLLINKRLKQGSVPDPSGM 687
Db      356 SETHQPMTSGSPFINGLHSHSE--VLTPARPOSMENSPFINE-----PKD----- 397

QY      688 LGLKVGGKMTESGRLCAPITKVKKGLADTVGHL-----RPGDEVLEWNGRLLQCAT 740
Db      398 -----KSELITLESVLPR-----SPADELSHCVEPEPSVPGGSSRDQ-----QGS- 440

QY      741 FEVYNIILSKPEQVQLVYSRPIGDIPRDPDTHAQLESS-----SSFSQKMDR-- 793
Db      441 -EEEPSVLKTLERSAARKMPKSLSDISS-DSSNQAKVDNQPEHIVLSAEDEKPDQKP 498

QY      794 -----PSISV--TSPMSP-----GMLRDVPQFLSQQLS---IKLWFKVGH----- 829
Db      499 VTNECVPRISVTPTOPDNFFPHDPDKLRMSKSVPAFLQDEVSGSVMSYSGDFGNLEVKG 558

```

```

Qy 830 -----OLVITLIGAKOLPSREGRPR-NPYVKIYFLPDRSDGKGRKTKTVKK 875
Db 559 NIQPAIEYVESLKLHRYFVAQCRLAAADYVKQQRSDFYVAYTLFPDKGRGKTKTVKK 618
Qy 876 TLSPKWNQTIYSPVHREERFEMLETTLDQARVREESSEFLGILLIETALLDDDEPH 935
Db 619 TLMFVYNEILRYK-IEKQILKTKNLSTVH--RDTYKRNSTFGLSEVLELDTWDMNDKQN 675
Qy 936 ----WYKLQTHD-----VSSLPLFRPSYLP 957
Db 676 KQLRWYPLKRTAPVALAENRGEMKALQVVPFVPGKLP 717

RESULT 7
US-10-144-198-2
; Sequence 2, Application US/10144198
; Publication No. US20030219748A1
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies Inc
; TITLE OF INVENTION: Regulated Prostate Cance Genes
; FILE REFERENCE: 9U 105 R1
; CURRENT APPLICATION NUMBER: US/10/144,198
; CURRENT FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 1564
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-144-198-2

Query Match 3.1%; Score 258.5; DB 12; Length 1564;
Best Local Similarity 18.9%; Pred. No. 7.8e-07;
Matches 286; Conservative 214; Mismatches 552; Indels 459; Gaps 73;

Qy 8 RGRPAFPA---ASQPPQPEMPDLSHLYTEKK---ILLAVMDRQKKREKESQSVLKIK 61
Db 203 RSKLSPSPSLRKSQSKSPKRSKPKSSAKDKRNTSAVSPFLLPQKNS-----KTN 254
Qy 62 REHKAQPTQWFPFSGITELVNNVLQPOKOPNEKEPOTLLHQPFFMYKQVKKMGESQQ 121
Db 255 QSKKGPRTPSPPPPPIED--ALGKKYK-----KYKVDRIEETKRDGKRGDRF 306
Qy 122 QOEQKADPATCGIKHTKPADGCHGNCYSQYTF---CARCGGRVSLRNNKVMVNCMLR 178
Db 307 QREKR-----DKPRSTSPAGQHSPISSRHHSSSSQSGSIQRHSPSPR-----R 351
Qy 179 KQEILTKSGANFWNSGNTLQPDQKVPGLRNEEAPQKKAKLHQPQPGAPGLSV 238
Db 352 KR-----TFPSY-----QRTLTTP-----LRASAFYPSHSL--SSPQKQSPPHRS 393
Qy 239 PAVEKGR-AHGLTKQDTIKNGSGVKHQAISMPDRKSPSVSRDQNRKYRQSEREDYS 297
Db 394 FWRKGRHDEHTSQSHDR-----RHEREDTRGRKREKDSREERYEQSGSSRDHRD 448
Qy 298 QYVESDGMFPRSPSDYADRSRQEPQYFEBHLNYRDSNNRGRHRSKRYIVDDDEYER 357
Db 449 DREPDGRDRDARDTRRELNDR-----DNEDS-----REMRDYSRDTK--ESR 493
Qy 358 DEYERQRRREHYQARYSDPNLAR-----YPVQPYTEQRIIHYVSRARHRRHSVSL 413
Db 494 DPRDSRSTRDAHYDRDEGRDTHKEPTYEESRSYG---RNHLRESSSTETAREERNE 550
Qy 414 ANALEDSRTLSLR-----MDPRQRQSVSERRAAMENQRSYEMERTREACQSSYPQR 467
Db 551 SRSEINRDMGRGRGVPELPEKSGRGRSGQIDSHSSNGNYHDSMETR-----SSYPR 605
Qy 468 TSNHSPPTPRSPILDRPDMRADSLRKQK-----HLDPSSAVR-----KTK 510
Db 606 DR-----YPER-----DNRQARDSSFERHGERDREDNRERQRPSPPIHQGNDELE 655
Qy 511 RKKMGTMLRNDLSLSQSGESVRPPPPPHKSKGKGRQVSLSSSESEELASTPEYTSDD 570

```


898 Db -----RKSERYEORLAKVSVDRQ-----RECTEILSSERMAQ---D 936
 826 Qy KVGHOLIVTILGANDLPSREDGRPNPVVLYFLPDRS-DNKRRTVTKVTLKPNQOT 884
 937 Db IIGH-----QSEDR-----TSDRAHDENKKAKIOKPIKKKKEDD 974
 885 Qy F-----IYSPVHRFRFRMLTLMQARVRESEBFLGILLIETAL 929
 975 Db VGIERNIETSEDOQVESP-KKGKKKSE-----KRRKSKG-----DSDI 1016
 930 Qy LDDEPHMYKQTHDVSLFLPPSPYLPRLQ-----HGESRTRLOKSKRISDSSEVDYD 985
 1017 Db SDEBAQOSKKRG-----PRTPITTKBELVEMCKNGKI--LEDSQKEDTAFSDWS 1068
 986 Qy CEDGVGVSDYRNGRDLQSLTSVPE--QVMSNHCSPPSGPHRVDVIGETB-WSPSA 1042
 1069 Db DE-----VFORTEVTEAHTATATTP-----GETPSPLSLL 1101
 1043 Qy PPPQR-----NVQGHGRTATGCHINTISMDRHR- 1072
 1102 Db PPPPVATATTPATLAATAAATSPSTSAITISTATPTNTN--NIFANEDSERK 1159
 1073 Qy VMDHYSDRDRCDRAADROPYHRSSTGEORPILLERTTRSRSSSRPDNLM 1124
 1160 Db CHRTRVEKVEPTHVIE-----DAHRKPMQKSSSS--LGSNRNRSHTSGR-----L 1206
 1125 Qy RSMPLMTGRSAPPSPAL-----SRSHPTGVSQVSPSTPTGRRGRLOLP-----P 1173
 1207 Db RSPNSDRAHSGDQSGKRVHLHSGRDRETKSLI-----TGERKSIDQKKEP 1259
 1174 Qy PKGT-----LERSAMD-----IEERNOMKLNKYQVAGSDPRLQDY 1211
 1260 Db SRSTSSDRQSRSSSRSSPESDQVHSRSGSFDRLQERDYEH-----DRERER 1315
 1212 Qy HSKYRSGWPHRGADTVSTKSSSDSDVSVASVSTSSASR---PSSSYMSVQSRPGRN 1268
 1316 Db RDTQREWD--RDADKOWPRNRDRERERERERERERERERERERERERERERERERER 1373
 1269 Qy RKISVFTS---KQNRQMGVSGKNLTKSTISGDMCSLEKNDGSQSTAVGAL-----GT 1320
 1374 Db DRDRTFSSQIESVKRCCKLEGEHERDLESTSRDLSALDK---ERMDKILGSGVGFEDT 1430
 1321 Qy SGKRRSSIGA 1331
 1431 Db NKSERTSLEA 1441

RESULT 9

US-09-987-482-1
 ; Sequence 1, Application US/09987482
 ; Publication No. US20020184656A1
 ; GENERAL INFORMATION:
 ; APPLICANT: BHANDARI, POONAM
 ; APPLICANT: SHASHIDHARA, L.S.
 ; TITLE OF INVENTION: IN VIVO ASSAY SYSTEM FOR SCREENING AND VALIDATION OF
 ; TITLE OF INVENTION: DRUGS AND OTHER SUBSTANCES
 ; FILE REFERENCE: 056859-0134
 ; CURRENT APPLICATION NUMBER: US/09/987,482
 ; CURRENT FILING DATE: 2002-03-21
 ; NUMBER OF SEQ ID NOS: 3
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 1
 ; LENGTH: 2843
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-987-482-1

Query Match 3.1%; Score 258.5; DB 10; Length 2843;
 Best Local Similarity 17.9%; Fred. No. 1.7e-06;
 Matches 304; Conservative 238; Mismatches 618; Indels 535; Gaps 69;
 30 SHLTERRKILAVMDRQKBEKEQSVLKIKIEHKAQTPQWPPSGITELVNNVLPQQ 89

1012 Db NEMDNDGELDTINTSLKTSDEQ-----LNSGRQSPSONERWARPKEILIE--DEIKOSEQ 1065
 90 Qy KOPNEKEPQKLHQQPMYKQSVKNGERSQO---QOSQKGDAPTGCICHTKTFADQCGH 146
 1066 Db QSRNQS-----TTTPTVTESTDDKHLKQPHHQOZ-----CVSPTRSRKANGSET 1112
 147 Qy NCSYCUTKFCARCGRVSLASNKVMVVCNLCKIQEILATKSGANFTNSGNTLQQDPQKV 206
 1113 Db N-----RVGNSHGINQVQSILC-----QEDDYEDDKPTNYSERVSEERQEE-- 1155
 207 Qy PRGLRNERAPOEKAKLHBOQPOQCAPGLSVPAVEKRAHGLTKQOTIKNGSGVHQHIA 266
 1156 Db EERTPNYSIKYNEEKHVDPQIDYSL-----QEDDYEDDKPTNYSERVSEERQEE-- 1155
 267 Qy SDMPDRKSPSVSRDQNRREYBOSEREDYSQVPSDGTMPRSPSDVADRSOREQPYE 326
 1185 Db TDIPSSQKQFSPSKSSQOQSKTE-----HMSSESTSTPSSNAKQOMLHPESAQ 1237
 327 Qy EPQELAYRDSNRGRHRSKE-----YIVDD-----EDVE 355
 1238 Db SRSQPOKAATCKVSSINQETIQTYCVEDTICFRCSSLSLSAARDEIGCNOTTQED 1297
 356 Qy SRDEYERQREHYQARYSDPNLARYPQPTBQORIHAEVRAKHERHSDVLAN 415
 1298 Db SANTLQIAETIKKIGITSAEDPVSEPAVSQHPRTKSRSLQGS-SLSSEARHKAVERFS 1356
 416 Qy AELEDSRISLLEMDRP-----SRQSVSERRAAMENORBYSMERTREAO--- 459
 1357 Db GAKSPSGAQTPKSPPEHVQETPLMFSRCTSVS-----SIDSFESSIASVSSQPCSG 1412
 460 Qy CQSSYPORTSNHSPPTFRSPIDLRDPMRADSLRKQHLDPSSAVKTKREK 513
 1413 Db MVSGIISPLDLPSPQTMPSRSKTPPP-----PPQTA--QTKRE- 1451
 514 Qy METWLRNDSLSQOSESVRPPPRPHKSKGKGRQVLSLSSSEEL-----ASTFETVTC 568
 1452 Db VPKNPTAKTKRES-----GPKQAVNAAVQVQLPFDADTLIHFATESTPDGFC 1502
 569 Qy D-----DVEL-----SSESVSEKDSQ-KGRKTI--SEOG 595
 1503 Db SSSLSALSLEDPFTQKDVLELRIMPVQENONGNETSESPKENGNEKCAKTIDSEK 1562
 596 Qy VLSDN-----TRSERQKRMVYGHSLSEDLSEWSEFQIKDSQVDTCS 639
 1563 Db LLDSDDDDIIELEECIISAMPTKSKKAKPAQASKLPPVAKPSQL-----PVYK 1616
 640 Qy TTLNEHSHSKHPVTWQPSKDGRL--IGRILIN-----KMLKQGVPRDSGAMGL 690
 1617 Db LLPSQNRLOPKH-VSPTPGDDMPRVYCVBCTPINFSTATSLSLDTTSPENELAGBGV 1675
 691 Qy KVGKMTESGRLCAPITKVKIGSLADTVGHLRPGDVLWNGRLLQCATPEEYVNLLE 750
 1676 Db R--GG--AQSG-----EFKRDITIPEG--RSTDEA-----QGGKTSVTIPELD 1714
 751 Qy SKRPEQVELV-----SRPICD-----IPRIDPSTHAQLESSSSSFESQMDPSISVT 799
 1715 Db DNKAEGDILABCSINAMPKSKHSHKPRVKINDQVQ-QASASSSAFKNQLDGKKKPT 1773
 800 Qy SPMSQMLRDPVQPLSQSLIKLWFKVGHQLIVTILGAKDLPSRDEGRPNPVKLYFL 859
 1774 Db SPVXP-----IPQWTEYRTRV-----KKVADSKNNLNAERVF 1805
 860 Qy PDRSKKRRKTKVKKLEPK-----MNQTFIYSPVH-----RR 893
 1806 Db SDNKSCKQNLKNSKDFNDKLPNNEDVRGSPAFDPSFHHYTFTEGTFYCFRSDNLSGL 1865
 894 Qy EFRMLEITLMDQARVRESEBFLGILLIETALLDDEPHMYKQTHDVSSLEPLR-- 951
 1866 Db DFDDDDVDLER-EKALRKAKENKESAKVTSHTELTSNQSANKTQA--IAKQPINRGQ 1922
 952 Qy PPSYLPRLQ-----HGESPTRRLQ-----SKRISDSSEVDYDCE--- 987

Db 1923 PKILOKSTPPOSSQIDIPRGAATDEKLNQFAIENTPVCFSHNSLSLSLSDIDQENNNK 1982
Qy 988 -----DGVGVSDVRHG-----RDLSSTLSPVEQVMSSNHC 1020
Db 1983 ENBFIKTEPPDSQESKQASGAVKSFVETTPVCFPSNLSLSLSDIDDLQEC 2042
Qy 1021 SPGSGPHRV-----VIGTRMSPPAPPORNVQEGHRTATGHYNTISMD-RHRVMD 1076
Db 2043 ISSAMPKKKPSRLKGNKXSP-----RNM-GGILGEDLTLDKDIQRPDSHGLSPD 2095
Qy 1077 HYSDDRDCEAD-----RQ-----PHERSTR 1101
Db 2096 SENFDWKAIOGANSIVSLSHQAACLSQASDSDSLKSLKSGISLSPFHLTPDQE 2155
Qy 1102 QRPLL-----ERTTKSSSRPDTN----- 1122
Db 2156 EKPTSNKGPRLKPGKSTLTKIESKIGKGGKVKYKSLITGKVSNSLSGQMKQ 2215
Qy 1123 -LMSMPSLMTGRAPPSPALSRHPTGVS-----QTSPTSPTGRRGRQLPOLPP 1174
Db 2216 PLQANMPSISRGRTMIHIPGVNSSSTSPVSKKGPPLKTPASKSPSEG-----QTATTP 2271
Qy 1175 KGTILERSAMDIERNRQMLNKYQVAGSDPLQDYHSKVRG-----WDPERGAD----- 1226
Db 2272 RGAKPSVKSELSVAROT-----SQIGSS-----KAPRSGRDSTPSRPAQPLSR 2319
Qy 1227 TVSTKSSDSVDVSASVRSSTASGRFSSTYSMSVQSERPRGNKISVFTSMQNRGMVS 1286
Db 2320 PIQPGNRSISPGENGISPPNKLQPLPTSPSTASTKSSGKMS-YTS--PGRQW--S 2374
Qy 1287 GKNITKTSISGDMCSEKNGDSQDTAVGALGTSGKGRSSIGAKMVAIVGLSRKNSA 1346
Db 2375 QONLTQGLSKNASSIPRSASAKGLNQMNNGANKK-----VELSRMSSTK 2423
Qy 1347 SOLSQTEGGKKL-----RSTVQRSTETGLAVENNMWMTQASREST----- 1388
Db 2424 SSGSESRSRPLVLRQFTFKEAPSTLRKLKESASFESLSPSRPASPTRSQATPV 2483
Qy 1389 -----DGMNSYSS 1397
Db 2484 LSPSLPDMSLSTHSS 2498

RESULT 10

US-08-681-219-32
; Sequence 32, Application US/08681219
; Publication No. US20020058607A1
; GENERAL INFORMATION:
; APPLICANT: Takaaki Sato and Junn Yanagisawa
; TITLE OF INVENTION: COMPOUNDS THAT INHIBIT THE INTERACTION BETWEEN
; TITLE OF INVENTION: SIGNAL-TRANSDUCING PROTEINS AND THE GLGF
; TITLE OF INVENTION: (PDZ/DHR) DOMAIN AND USES THEREOF
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESS: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/681,219
; FILING DATE: 22-JUL-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 0575/48962/JPM/JM

TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2843 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-681-219-32

Query Match 3.1%; Score 257.5; DB 8; Length 2843;
Best Local Similarity 18.0%; Pred. No. 2e-06;
Matches 305; Conservative 238; Mismatches 617; Indels 535; Gaps 70;

Qy 30 SHLTSEERKILAVMDRQKKKEKESQSLVKIKEHKAQPTQWPFPSGITELVNVLPQOQ 89
Db 1012 NHDDNDGELDTINSLKYSDEQ-----LNSGRQSPQNEKMARPKHIL-DEIKQSEQ 1065
Qy 90 KQNEKEPQTKLHQFEMKYQVKMGHEEQ---QOEKGDAPTGGICHKTKFADCGH 146
Db 1066 RQSRNQS-----TTPVYTESDDKHLKQPHFGQOE-----CVSPYRSRGANGSET 1112
Qy 147 NCYQCTKPCARGVSLSNKVMVNCILCRQOEILLITSGANFYNSGNTLQOPQKV 206
Db 1113 N-----RVSGNHGINQVQSILC-----QEDDYEDDKFTNYSRYSEESQHEE-- 1155
Qy 207 PRGLNEEAQEKKAKLHQPOFGAPGLDLSVPAVEKRAHGLTRQDTIKNGSOVKHQA 266
Db 1156 -----EERTNYSIKNEKRHVDPIDYSI-----LK-----A 1184
Qy 267 SDMPDRKSPSVSRDQRRYQSEEREDYSQVYVPSGTMTSPSPDYADRRSQREPFYE 326
Db 1185 TDIPSSQKQSFSSKSSSQSSKTE-----HMSSESTSTPSSNAKQNLHPSSAQ 1237
Qy 327 EPGLHLYDNRGRHRSKE---YIVDD-----EDVE 355
Db 1238 SRGQPKAATCKVSSINQITQICVEDTPICFRCSLSLSLSSASDEICGNTTQDQ 1297
Qy 356 SRDEYRQRREHYQARYRSDPNLARYPVKQFPQVEHQRIHARYSRHRRHSDVSLAN 415
Db 1298 SANTLOIAEIKELIGTASAEDEPVSEVPAVCHPTKSRQGS-SLSESAKHAVFSS 1356
Qy 416 AELDSRISLRLMDRP-----SRQSVSEERRAAMENQSYSMETRAQ----- 459
Db 1357 GAKSPSKGATPKSPPEHYVQETPLMFSCRCTSVS-----SLDSPESRSLASSVQSEPCSG 1412
Qy 460 -----QSSVYPORTSNHSPTPERSPIPLDRPDMRADSLKQHLDPSSAVKTKREK 513
Db 1413 MVSGIISPSDLFDSFGQTMPPSPRSKTPPP-----PPQTA---QTKRE- 1451
Qy 514 METMLRNDLSLSDQSESVRPPPHPSKKGKQKQVLSLSSREEL-----ASTPEYTS 568
Db 1452 ---VPKNAPTAKRES-----GPKQAAVNAVQVQVLPDADTLHPATSTPDGFC 1502
Qy 569 D-----DVEL-----ESSEVSEKDSQ-KGRKKT---SEQ 595
Db 1503 SSSLALSLEPFTQKQVLAIMPVQENDNGNETESEQPKESNENQKRAKTIIDSEKD 1562
Qy 596 VLSDSN-----TSEROKKEMTYGHSLEEDLEWSPOIKDSQVDTCS 639
Db 1563 LLDDSDDDDIIEECIISAMPTKSSRAKKPAQATSKLPPVAKPSQL-----PVYK 1616
Qy 640 TTLNEHSHSKHPVTWQPSKDGRL--IGRILN-----KLLKDGSVPRDSGAMGL 690
Db 1617 LLPQNRLOPKH-VSTPGDMRVYCVETGPIFSTATSLDITTESPWEAAGEV 1675
Qy 691 KVGGMKTESGRLCATFKVKKGLADTVGHLRPGDEVLEWNGLLQATPEEYVNIIL 750
Db 1676 R--GG--AQSG-----EFEKRDITPTEG--RSTDEA-----QGGKTSVTIPELD 1714
Qy 751 SKPPEQVELV-----SRPIGD-----IPRIPDSTHAQLESSSSSESQMDRPSISVT 799

```
Db 1715 DNKAEEDGILLACINSAAPKSHKPRVKKIMDQVQ-QASASSAPAKNQLDCKKKPT 1773
Qy 800 SPMSPGMLRDVPOPLSQSLKLFKDFKGVHQLIVITILGAKDLPSPRGPRPNVVKIYFL 859
Db 1774 SPVKP-----IPQNTYTRTV-----RKNADSKNNLNAERVF 1805
Qy 860 PDRSDKMRRTKTKVKTLEPK-----WNOTFIYSPVH-----RR 893
Db 1806 SDNKSQKNNLKNNSKOPNDKLPNNEDVRGSPAFDSPHVTPLEGTPYCFSRNDSLSSL 1865
Qy 894 EFRERMLEITWDOARVRESESEFLGELILEFTALLDDDEPHWTKLQTHDVSSLPFR-- 951
Db 1866 DFDDDDVDLSR-EKAELEKAKENKSEAKVTSHTLSAQSNKQQA--IAKQPINRGQ 1922
Qy 952 PSLVLPRLQ-----HGEPTRLQR-----SKRISSEVSDYCE-- 987
Db 1923 PKPILQKQSTPPQSQKIDPRGAATDEKLNFAIENTFVCFESHSSLSLSDIDQENNK 1982
Qy 988 -----DGVGVSDYRHNG-----RDLQSSTLVPEQVMSSNHC 1020
Db 1983 ENBPKEKTEPPDSQGEPSKPAQSGVAPKSPHVEDTPVCFSENSSLSLSDIDQENNK 2042
Qy 1021 SPGSPHRVD---VIGTRSWSPAPPORNVQOHRGTRATGHTWNTISMD-PHYVDD 1076
Db 2043 ISSAMPKKKPKPSRLKGNKHSK-----RNM-GGILGEDULDLKDIOQPDSEHGLSPD 2095
Qy 1077 HYSDDRDCEAAD-----RQ-----PYHRSRSTE 1101
Db 2096 SENFDWTKAQEGANSIVSSLLHQAACLSRQASDSDLSLSGSLGSPFHLTPDQE 2155
Qy 1102 QRLP-----ERTTRSRSPRPTN----- 1122
Db 2156 EKPTSNKGPRLKPGKSTLETKKISESKGKGVKYLITGVRSNSELGQWKQ 2215
Qy 1123 -LMSMSLMTGRAPSPALSRHPTGVS-----QTSPTSTGTRGRQLPOLPP 1174
Db 2216 PLOAMFPISEKRTWIIHPIGVNRSSTSPVSKKGPPLKTPASKSPSEG---QTATTSP 2271
Qy 1175 KGTILERSAMDIERNRQMKLNKYQVAGSDPRLQDYHVKYRG---WDPERGAD--- 1256
Db 2272 RGAKPSVKSELSFVARQT-----SQIGSS-----KAPSRGSRDSTPRPAQPLSR 2319
Qy 1227 TVSTKSSDSDVSDVSAVSRSTASRFSSTSMVSQSERPRGNKISVFTSKQNRQMGVS 1286
Db 2320 PIQSPGNSISPRNGISPPNKISQLPRTSPSTASTKSSGKMS-YTS--PGKQW--S 2374
Qy 1287 GKMLTKSTSISGDMCSLEKNDGSDTAVGALGTSGKKRSISGAKWVAIVLSKRSRA 1346
Db 2375 QONLTKOTGLSKNASSIPRSASAKGLNQMNGANKK-----VELSRMSSTK 2423
Qy 1347 SOLSQTEGGKKL-----RSTVORSTETGLAVENNMWTRQASREST----- 1388
Db 2424 SSGSESRSERPVLVQSTPFIKEAPSPILRKLRESASFESLSPSRPASPTRQAQTFV 2483
Qy 1389 -----DGMNSYS 1397
Db 2484 LSPSLPDMSLSTHSS 2498
```

RESULT 11

US-10-092-138-30

; Sequence 30, Application US/10092138

; Publication No. US20030170723A1

; GENERAL INFORMATION:

; APPLICANT: Sato, Taka-Aki

; TITLE OF INVENTION: METHOD OF PREPARING A PROTEIN ARRAY BASED ON

; FILE REFERENCE: BIOCHEMICAL PROTEIN-PROTEIN INTERACTION

; CURRENT APPLICATION NUMBER: US/10/092,138

; CURRENT FILING DATE: 2002-09-06

; NUMBER OF SEQ ID NOS: 33

; SOFTWARE: PatentIn Ver. 2.1

```
; SEQ ID NO 30
; LENGTH: 2843
; TYPE: PRT
; ORGANISM: human
; US-10-092-138-30
```

Query Match

3.1%; Score 257.5; DB 12; Length 2843;

Best Local Similarity 18.0%; Pred. No. 2e-06;

Matches 305; Conservative 238; Mismatches 617; Indels 535; Gaps 70;

```
Qy 30 SHTLEERKILLAVMDRQKEEKEQSVLKIKKEHKAQPTQWFFPSGITLVNVLQPOQ 89
Db 1012 NMDDNDGELDTPINTSLKTSDEQ-----LMSGQSPSQNERWARPKHILB--DEIKQSE 1065
Qy 90 KQWKEKEPOTKLHQCFEMAYEQVKKGHESSQ---QOQKGDAPTCIGICHTKTFADQCGH 146
Db 1066 RQSENQS-----TTPVVTSTDDKHLKQPHFQOE-----CVSPTRSGANSET 1112
Qy 147 NCYCYQTKFCARCGRVSLSNKVMVNCNCRQOEBILTKSGAWFYNSGNTLQQPDQKV 206
Db 1113 N-----RVGSHNGINQNVQSILC-----QEDDYEDDKPTNYSERYSEERQEE-- 1155
Qy 207 PRGLRNEEAPQEKKAKLHQPOFGAGDLSVPAVEKRAHGLTROPTIKGSGVKGHQA 266
Db 1156 -----EERTNTSIKINEKRRHVDQIDYSI-----LK-----A 1184
Qy 267 SDMPDKRSPSVSRDQNRREYQSEEREDYSQTVPSDGTMPRSPSDYADRSQREPOFYE 326
Db 1185 TDIPSSQKQSPFSKSSGQSSKTE-----HMSSENTSTPSSNAKQKQHLFSSAQ 1237
Qy 327 EPGLHYTRDNRGRHRSKE-----YIVDD-----BDVE 355
Db 1238 SRSQCPQKAAATCKVSSINQETIQTYCVEDTPICFRCSLSLSSAEDHETGCTTQEAD 1297
Qy 356 SRDEYERQRESEYQARYRSDPNLARVPKQPVQVEQRIHAFVSRARHRRSDVSLAN 415
Db 1298 SAMLQIAEIKKIGKIGTSAEDPVSEVPASVQHPRTKSSRLQGS-SLSESARHKAVERFS 1356
Qy 416 AELEDSRISILLRDRP-----SRQSVSEERRAAMENQSYSEMERTEAQ--- 459
Db 1357 GAKSPKSGAQTPKSPPEHYVQETPLMFSRCTSVS-----SLDSFESBIASSVQSEFCSG 1412
Qy 460 -----QSSYQRTSNHSPPTPRRSPILDRPDMRADSLRKQHHLDPSSAVKTKREK 513
Db 1413 MVSGIILSPDLPSQGTMPFSRKTTPP-----PPQTA--QTKRE- 1451
Qy 514 METWLRNDLSLSDQSESVRPPPRPKSKKGGKQKQVYLSLSEEL-----ASTPEYTC 568
Db 1452 ---VPKNKAPTAKRES-----GPKQAAVNAVQVQVLPDADTLILHFAETSTPDGFSC 1502
Qy 569 D-----DVEL-----ESESVESEKDSQ-KGKRTK--SEQG 595
Db 1503 SSSLALSLEDPFIQKQVLELRIMPPVQENDNGNETSEEQPKESNENQEKAEKTIIDSEK 1562
Qy 596 VLSDSN-----TRSERQKQKMYTGGHSLHEDLSEWSFQKSDGVTCSS 639
Db 1563 LLDDSDDDDIIELEECIIISAMPTKSRKAKKPACTAKLPPVARKPSQL-----PVYK 1616
Qy 640 TTLNEESHSHDKHPVTWQPSKQDGL--IGRIILN-----KELKQGSVPRDSGMGL 690
Db 1617 LLPQNLQPKQH-VSPITGDMRVCVSGTPINFSTATSLDLTSTSPNLEAGEV 1675
Qy 691 KVVGKRMTEGRLCAPITKVKKGLADTVCHLAPGDEVLWNGRLKLGAFVRYNILE 750
Db 1676 R--GG--AQSG-----EFKRTDTIPEG--RSTDEA-----QCKTSSTVIFELD 1714
Qy 751 SKPEQVELV-----SRPTG-----IPRIPDSTHAQLESSESSSESQMDRPSISVT 799
Db 1715 DNKAEGDILLAEICINSAMPKSKHPRVKKIMDQVQ-QASASSAPAKNQLDCKKKPT 1773
Qy 800 SPMSPGMLRDVPOPLSQSLKLFKDFKGVHQLIVITILGAKDLPSPRGPRPNVVKIYFL 859
Db 1774 SPVKP-----IPQNTYTRTV-----RKNADSKNNLNAERVF 1805
```

```

QY 860 PDSKDKRRTKTKVKTLEPK-----WQTFIYSPVH-----RR 893
Db 1806 SDNKDSKQKLNKNSKDFNDKLPNNEDVRGSPAFDPSPHYTPIEGTPTCYFSRNDLSL 1865
QY 894 EFERMLEITLWQARVRESEFLGELIILETALDDDEPHYKQLQTHDVSLPLPR-- 951
Db 1866 DFDDDDVILSR-EKALEKAKENKESEAKVTSHTLTNSQOSANKTQA--IAKQINROQ 1922
QY 952 PSLYPRQL-----HGSEPTRLQR-----SKRIGDSKVSVDYDCE-- 987
Db 1923 EKILQKQSTFPOSSKDIDPRGAATKELQNFALIENTPVCFSINSSLSLSDIDQENNK 1982
QY 988 -----DGVGVVDYRHNG-----RDLQSTLSVPEQWMSNEC 1020
Db 1983 ENBPKEPTEPDQGRPKPQAGYAPKSPHVEDTPVCFGRNSSLSSLSIDSDLLQEC 2042
QY 1021 SPGSPHRVD---VIGETRWSAPSAPPQBNVQCHRGHTATGYNTISMD-RHRVMD 1076
Db 2043 ISSAMPKCKKPSRLKGNKHSK-----RNM-GGILGEDLTLDLKDIOQPDSEHGLSPD 2095
QY 1077 HYSDDRDCEAD-----RQ-----PYHRSRST 1101
Db 2096 SENFOWKAIQEGANSIVSSLHQAAAACLSRQASSDSLSLSKSGISLSPHLPDQ 2155
QY 1102 QRPILL-----ERTTRSRSSRPDTN----- 1122
Db 2156 EKPFSTNKGPRILKPGKSTLETKIESKSGIKGKVKYKSLITGKVRNSEISQMKQ 2215
QY 1123 -LWRSPLMTGHSAPSPALSHPTGVS-----QTSPTPTGTCRGQLQLPP 1174
Db 2216 PLQAMPISIRGRTHIHPGRVNSSSTSPVSKKGPPLKTPASKSPBG---QTATTSP 2271
QY 1175 KGTLSAMDIEERNQMLKMKYQVAGSDPRLEQDVHSYRSG---WDPHRGAD-- 1226
Db 2272 RGAKPSKSLSPVAKT-----SQIGSS-----KAPSRGSRDSTSRPAQPLSR 2319
QY 1227 TVSTKSDSDVDSVAVSTSSASRFSSTYSMSVQSRPRGNRKISVFTSKMQRNGVS 1286
Db 2320 PIQSPGRNSISPRNGISPPNKISQLPRTSPSTASTKSSGSGKMS-YTS--PGQW--S 2374
QY 1287 GKULTKSTISGDMCSLEKNDGQSDTAVGALGTSKGRKSSIGAKVAIVGLSRKRS 1346
Db 2375 QOBLTKQTLGSLKXNASSIPRSESASKGLNOMNNGANKK-----VELSRMSTK 2423
QY 1347 SOLSOTRGCKKL-----RSTVORSTETGLAVEMNMTQASREST----- 1388
Db 2424 SSGESDRSRPVLVROSTFIKEAPSTILRKUEESASFESLSPSRPASPTRSQACTPV 2483
QY 1389 -----DGSNMNYS 1397
Db 2484 LSPSLDMSLSTHSS 2498

```

RESULT 12

```

US-09-230-111C-30
; Sequence 30, Application US/09230111C
; Publication No. US20030203414A1
; GENERAL INFORMATION:
; APPLICANT: Sato, Taka-Aki
; APPLICANT: Yanagisawa, Junn
; TITLE OF INVENTION: COMPOUNDS THAT INHIBIT INTERACTION BETWEEN
; TITLE OF INVENTION: SIGNAL-TRANSDUCING PROTEINS AND THE GLGF (PDZ/DHR)
; TITLE OF INVENTION: DOMAIN AND USES THEREOF
; FILE REFERENCE: 48962-A-PCT-US
; CURRENT APPLICATION NUMBER: US/09/230.111C
; CURRENT FILING DATE: 1999-05-17
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 30
; LENGTH: 2843
; TYPE: PRT
; ORGANISM: human

```

US-09-230-111C-30

```

Query Match 3.1%; Score 257.5; DB 12; Length 2843;
Best Local Similarity 18.0%; Pred. No. 2e-06;
Matches 305; Conservative 238; Mismatches 617; Indels 535; Gaps 70;

QY 30 SHLTREERKILIAVMDRQKKEEKSQVLKKEHKAQTPWPPSGITELVNVLOPOQ 89
Db 1012 NMDNDGELDPINISLKYSDQ-----LNSGRQSPQSRWARPKHIE--DETQSSQ 1065
QY 90 KQPKNEKPTQLHQOFEMYKQVKMGESQ---QOEQKGDAPTGCIGCHTKTFADGCH 146
Db 1066 RGRNQS-----TTPVTTESTDDKHLKQPHFQOE-----CVSPYRGANGST 1112
QY 147 NCSYQOTKFCARGGVRSLRNKVMVNCLEKQOEILITKSGAMFNSSMTLQPDQKV 206
Db 1113 N-----RVGSHNGINQVQSQC-----QEDDYEDDKPTMYSERVSEKHEE-- 1155
QY 207 PRGLRNEEAPQEKAKLHPOPOQAPGDLSPAVEKGRAGLITRODTTKNGSGVHQIA 266
Db 1156 -----EERTMYSTKYNEEKHVPQIDYSI-----LK-----A 1184
QY 267 SDMPDREKSPSVSRDONRRYQSBEREDYSOVVPSDGTMPRSPDYADRRSQRPOFYE 326
Db 1185 TILPSSQKQSPFSKSSSQSKT-----HMSSESTSTPSSNAKQOLHPSSAQ 1237
QY 327 EPGHLNYSRNRGRHRSXE-----YIVD-----EDVE 355
Db 1238 SRSGQPKAATCKVSSINQETIQTVCVEDTIPCSPRCSLSLSAEDHICGNTTQEAD 1297
QY 356 SRDEYRQREHEQARYSDPNLARYPVQPOPYEQMHIHARVSRARHRRHSVSLAN 415
Db 1298 SANTLQIAIEKIKIGTSAEDFVSEVPAVSQPRTKSSLOQS-SLSSSRARHKAVERFS 1356
QY 416 AELEDERSILLNRDRP-----SKQSVSERRAAMENQRSYMERTEAQ----- 459
Db 1357 GAKSPKSGAQTSPKSPHYVQETPLAFSRCTSVS-----SLDSFERSIASVQSPSCG 1412
QY 460 -----GQSSYQRTSNHSPPTPRRSPIFLDRPDMRRADSLRQHLDPSSAVKTKRK 513
Db 1413 MVSGIISPSDLPDSQPTWPPSKTPTP-----PQTA--QTKRE-- 1451
QY 514 METMLANDSLSDQSVSRVPPPPRPHKSKKGMROVSLSSSEEL-----ASTPYTSC 568
Db 1452 ---VPKNKPTAKRES-----GPKQAAVNAVQVQVLPDADTLHPATESTPDFSC 1502
QY 569 D-----DVEL-----ESRSVSKGDSQ-KGKRT--SEQ 595
Db 1503 SSSLALSILDEPPIQKDVLRIMPVQENDNGNETESQPKSENENQEKAEKTIIDSKD 1562
QY 596 VLSDSN-----TRSERQKRMYYGHSLEEDLEWSEPOKDSGVDTCS 639
Db 1563 LLDSDDDDDIEILEECIISAMPTKSKAKQAQATASKLPPPVARKPSQL-----PVYK 1616
QY 640 TTLNEHSHSDKHPVTWQPSKGDRL--IGRILLN-----KLLKQGVPRDSGMLGL 690
Db 1617 LLPSONRLQPKH-VSFTPCDDMPRYCVGEGTINFSTATSLDLTIESPPNELAAGEV 1675
QY 691 KVVGGNMTESGRICAFITKVKGSLADTVGHLRPGDVELEWNGRLQCATFREVYVILR 750
Db 1676 R--GG--AQSG-----EFKDDTPTREG--RSTDEA-----QGGKTSVTPIELD 1714
QY 751 SKPEPQVELVW-----SRPIGD-----IPRIDPSTHAQLESSESSSFBZKQDRPSIVT 799
Db 1715 DNKAESGDILAEICNSAMPKSKHPRVKIWDVQV-QASASSAPAKNQLDGGKKKT 1773
QY 800 SPMSPGMLNDVQFLSGQLSILKMFQKVGHLIVTILGAKDLPSSDRGRPRPVKIYFL 859
Db 1774 SPVKP-----IPQNTYRTRV-----RKNADSKNLNAERVF 1805
QY 860 PDSKDKRRTKTKVKTLEPK-----WQTFIYSPVH-----RR 893
Db 1806 SDNKDSKQKLNKNSKDFNDKLPNNEDVRGSPAFDPSPHYTPIEGTPTCYFSRNDLSL 1865

```

894	QY	EFRMLIEITLWDOARVRREBESFLGEILIELETALLDDEPHWYKLOTHDVSSLPFR--	951
1866	Db	DFODDDYDLGR-EKAEILRKAKENESAKVTSTHTLSNQOASAKTQA--IAKQPIKRGQ	1922
952	QY	PSFVLPBRQL-----HGESPTRELQR-----SKRISDSVSDYDCE-----	987
1923	Db	PKPILQKQSTPQSSKIDPRGAATDEKQLQFALIENTPVCFSHNSLSUSSDIDQENNK	1982
988	QY	-----DGVGVSVDYRHNG-----RDLQSSSTLSVPEQWSSNNHC	1020
1983	Db	ENRPIKETEPDQSCQEPKQASGYAPKSFHVEDTPVCFSRNSSLSSLSIDSEDDLQEC	2042
1021	QY	SPSGSERVD---VIGTRSWSPCAPPPQNRVEQHGRTATGHYNTISNWD--RHRVDD	1076
2043	Db	ISSAMPKKKPSRLKGNEXKHP-----RNM--GGIIGEDLTLDLKDIORPDSHGLSPD	2095
1077	QY	HYSDRDRCBAAD-----RQ-----FYHRSRSTE	1101
2096	Db	SENFDMKAIQEGANSIVSSLHQAAAAACLGRQASSDSILSLKSGISLGSPFHLTPDQE	2155
1102	QY	ORPLL-----ERTTTRSSSRPDTN-----	1122
2156	Db	EKPFTSNKGPRILLKPGSKSTLTKKIESESGKIGKKVKYKSLITGKVRNSSEISGQWKQ	2215
1123	QY	LAKRSMPLATGRSAPPSPALSRHPRTGSV-----QTSPSSTPGTGRGRQLPQLPP	1174
2216	Db	PLQANMPSISRGRTMIHIGVRNSSSSTSPVSKKGPPLKTPASKSPSEG-----QTATSP	2271
1175	QY	XGTLERSAMDIIEENRQMKLNKYQVAGSDPRLEQDVHSKYRSG-----WDPHRGAD-----	1226
2272	Db	RGAKPSVKSELSFVARQT-----SQIGSS-----KAPSRSGRSTSPRPAQPLSR	2319
1227	QY	TVSTKSDSDVDVSASVRSSTSSARPSSTSTVMSVQSERPRGNKRI SVFTSKMQRQMGVS	1286
2320	Db	PIQSPGENSISPGENGISPPNKLSQLPRTSSPSTASTKSGSGKMS-YTIS--PCROM--S	2374
1287	QY	GKNLTKSTSTSGDMCWSLEKNDGGQSDTAVGALGTSGKRRSSIGAKMVAIVGLSRKRSA	1346
2375	Db	QONLTKQTLGSKNASSIIPRESASKGLNOMNNGANKK-----VELSRMSYTK	2423
1347	QY	SQLSQTEGGGKCL-----RSTVORSTETGLAVENRWMTTQASREST-----	1388
2424	Db	SGSESDRPRVLVRQSTFIKEAPSTPLRRKLEESASFESLSPSSRPASPTTSSQATPV	2483
1389	QY	-----DGSNMNSYSS	1397
2484	Db	LSPSLPDMSLSTHSS	2498

RESULT 13

US-10-311-626-5
/ Sequence 5, Application US/10311626
/ Publication NO. US20030186379A1
/ GENERAL INFORMATION:
/ APPLICANT: INCYTE GENOMICS, INC.
/ APPLICANT: LAL, Preeti
/ APPLICANT: TANG, Y. Tom
/ APPLICANT: YUE, Henry
/ APPLICANT: CHAWLA, Narinder K.
/ APPLICANT: BAUGHN, Mariah R.
/ APPLICANT: DAS Debopriya
/ APPLICANT: RAMKUMAR Jeyalaxmi
/ APPLICANT: LIU, Dyung Aina M.
/ APPLICANT: HAFALIH, April
/ APPLICANT: GANDHI, Ameena R.
/ APPLICANT: LEE, Ernestine A.
/ APPLICANT: XU, Yuming
/ APPLICANT: BANDOMAN, Olga
/ APPLICANT: ELLIOT, Vicki S.
/ APPLICANT: NGUYEN, Damiel B.
/ APPLICANT: BURRILL John D.

APPLICANT: MARCUS, Gregory A.
APPLICANT: ZINGLER, Kurt A.
APPLICANT: LU, Yan
APPLICANT: YAO Monique G.
APPLICANT: GURURAJAN, Rajagopal
APPLICANT: DING, Li
APPLICANT: WARREN, Bridget A.
APPLICANT: THANGAVELU, Kavitha
APPLICANT: LEE, Sally
TITLE OF INVENTION: SECRETION AND TRAPPING MOLECULES
FILE REFERENCE: PF-0801 PCT
CURRENT APPLICATION NUMBER: US/10/311,626
CURRENT FILING DATE: 2002-12-17
PRIOR APPLICATION NUMBER: 60/215,465; 60/239,384; 60/253,639
PRIOR FILING DATE: 2000-06-29; 2000-10-10; 2000-11-28
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PERL Program
SEQ ID NO 5
LENGTH: 671
TYPE: RPT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. US20030186379A1 4728247CD1
US-10-311-626-5

```

Query Match      3.0%; Score 252; DB 12; Length 671;
Best Local Similarity 18.8%; Pred.No. 6.1e-07;
Matches 184; Conservative 117; Mismatches 278; Indels 402; Gaps 39;

QY      25  EMDLSHLTREERKILAVMDRCKEKEKESVLKIEEKAQTQWPPSGITELVNVV 84
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      3  ELDDLSPLEEKDLILLVLQR-----DEEVKADEKR-----IRRLNEL 43
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      85  LQPOQKPNKEKPTKLHQGFMYKEQVCKGHE--SQQEQEOKGDAPTCGLCH- 136
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      44  L-----EIKKKAEGSGHYDR-----TCARQESLRL 73
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      137 --KTKPADGGCHNCYCQTKFCACGRVSLRKNVMVNCILKQKQELITKSGAFYNS 194
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      74  SPKINTCRGCNH-----LVCRDC---KIQESNGTWRCNKVCAKEILKKAQSDMYF-- 120
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      195  GSNYLQPDQKVRPGLNEEAPQKAKLHQFPQGPAGDLSVPAVEKGRAGHLTRQDT 254
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      121  ---DQVNRFAYRTGSEITIRMSLRK-----PAVSK-----RET 151
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      255  IKNGSGVKHQI-----ASDMPSDRKRSPSV-----SEDQNRREYQSEEREDYQYV 300
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      152  V--GQSLLHQTQMGDIWPGKIIQERQKEPSVLFEVPLKSGSKSALESLSDFT--A 207
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      301  PSDGTWPR-----SPSDYADRSQREPO--FYEEPCHLYNRDSNRGHR 342
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      208  DSDSTSRRLSLDKSGLFPWKMSAPKSVQEKETPGQGNVVFVDEGEMFKKTKILR 267
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      343  HSEYIVDDDEVSRUEYERQREHEEQARYRSDENLARYVPQPYEQEQRHAEVSR 402
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      268  PS-EYTKSVTDLRPEDVWHESGSLGD---RASKSVPL-----300
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      403  RHERRHSDYSLANAEKLEDSIILRMDRPSQESVSERRAAMENQRS---YS----- 451
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      301  -----NVDMEBESEEDIDHLVTLHQKLARSSMQSGSEMTSGMSIYSEAGDFGN 353
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      452  MERTBAQOGSSYPQRTSN---HSPTFRSPITLDRPDWRADSLRKCHHLDPSSAVRK 508
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      354  IFYTGRIASLKEEQTSILVH-----VKECHLAYADRAKK 391
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      509  TQREKMETMLRNDLSLSDQSESVRPPPHKSKKGGKMKQVLSHSEELASTFEYTSC 568
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      392  RSNPYVKTLLDPK-----405
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      569  DVELESESVSEKDGQKGRKRTSBOGVLSDNTRSERQKRWYGGHSLSEEDLWSEPQ 628
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      406  -----SQGGKRTS-----IKRDTVNPVLY-----DETLEVEIPE 434
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

US-09-864-761-38199
Sequence 38199, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Aecmics-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408

Result No.	Query		DB	ID	Description
	Score	Match Length			
1	307	3.7	1898	1	US-08-056-200-94
2	307	3.7	1898	2	US-08-000-644-94
3	258.5	3.1	2842	1	US-07-741-940-7
4	258.5	3.1	2842	1	US-08-289-548A-7
5	258.5	3.1	2842	1	US-08-452-654-7
6	258.5	3.1	2842	4	US-08-449-731-7
7	258.5	3.1	2843	1	US-08-452-655B-2
8	258.5	3.1	2843	1	US-08-452-655B-7
9	258.5	3.1	2843	3	US-08-450-582-2
10	258.5	3.1	2843	3	US-08-450-582-7
11	258.5	3.1	2973	2	US-08-821-355A-7
12	258.5	3.1	2973	2	US-09-003-687A-7
13	258.5	3.1	2973	3	US-09-136-605-7
14	250.5	3.0	2843	1	US-07-741-940-2
15	250.5	3.0	2843	1	US-08-289-548A-2
16	250.5	3.0	2843	1	US-08-452-654-2
17	250.5	3.0	2843	2	US-08-370-235A-2
18	250.5	3.0	2843	4	US-08-449-731-2
19	234	2.8	1896	4	US-08-978-277A-4
20	227	2.7	8991	4	US-08-714-741-32
21	225.5	2.7	1780	1	US-08-769-309A-5
22	225.5	2.7	1780	3	US-08-994-570-5
23	222	2.7	2101	1	US-08-466-390-4
24	222	2.7	2101	1	US-08-470-950-4
25	222	2.7	2101	1	US-08-467-781-4
26	222	2.7	2101	2	US-08-483-924-4
27	222	2.7	2101	3	US-09-452-294-1
28	222	2.7	2101	3	US-09-452-294-1
29	222	2.7	2101	3	US-09-452-294-1
30	222	2.7	2101	3	US-09-452-294-1
31	222	2.7	2101	3	US-09-452-294-1
32	222	2.7	2101	3	US-09-452-294-1
33	222	2.7	2101	3	US-09-452-294-1
34	222	2.7	2101	3	US-09-452-294-1
35	222	2.7	2101	3	US-09-452-294-1
36	222	2.7	2101	3	US-09-452-294-1
37	222	2.7	2101	3	US-09-452-294-1
38	222	2.7	2101	3	US-09-452-294-1
39	222	2.7	2101	3	US-09-452-294-1
40	222	2.7	2101	3	US-09-452-294-1
41	222	2.7	2101	3	US-09-452-294-1
42	222	2.7	2101	3	US-09-452-294-1
43	222	2.7	2101	3	US-09-452-294-1
44	222	2.7	2101	3	US-09-452-294-1
45	222	2.7	2101	3	US-09-452-294-1
46	222	2.7	2101	3	US-09-452-294-1
47	222	2.7	2101	3	US-09-452-294-1
48	222	2.7	2101	3	US-09-452-294-1
49	222	2.7	2101	3	US-09-452-294-1
50	222	2.7	2101	3	US-09-452-294-1
51	222	2.7	2101	3	US-09-452-294-1
52	222	2.7	2101	3	US-09-452-294-1
53	222	2.7	2101	3	US-09-452-294-1
54	222	2.7	2101	3	US-09-452-294-1
55	222	2.7	2101	3	US-09-452-294-1
56	222	2.7	2101	3	US-09-452-294-1
57	222	2.7	2101	3	US-09-452-294-1
58	222	2.7	2101	3	US-09-452-294-1
59	222	2.7	2101	3	US-09-452-294-1
60	222	2.7	2101	3	US-09-452-294-1
61	222	2.7	2101	3	US-09-452-294-1
62	222	2.7	2101	3	US-09-452-294-1
63	222	2.7	2101	3	US-09-452-294-1
64	222	2.7	2101	3	US-09-452-294-


```

Db      1051 LREER-----EKRLQREKQYREBELQOREEQQLIGREKFTRRQELERQYRK 1099
QY      1113 SRSSREFDTNLRSMPSLMTGRSAPPSPALSRSHPTGVSQTSPTSPTGTRGRQLPOL 1172
Db      1100 EELQOREEQQLARE-----EPEKRRQERERQCRREBEELQOREEQQLREERK 1147
QY      1173 PPGTTLERAMDIERNRMKLNKYQVAGSDPRLBQDTHSKYTRSGWDPH----- 1222
Db      1148 RRQELERQYREBEELQORQKRIQRYA-----DEDQSRDLKQWPEPEKENAVRDNKV 1198
QY      1223 --RGADTVSTKS--SDSDVDVSVAVSTSSASRPSSTSYMSVQSERPRGNRKLSVFTSKMQ 1279
Db      1199 YCKRGREWFQRLHDSQVRD-----RQSOODLQHLIGEOERDR-----EOE 1240
QY      1280 NRQMGVSGKNLTKTSTISGDMCSLEKNDGQSQTAVGALGTSGKRSSSIGANVAIVGL 1339
Db      1241 RRMQQAHRHFPEE-----QLEREQEKEA-----TKR----- 1268
QY      1340 SRKRSASQLSOTFEGGKKLRSTVQR 1365
Db      1269 DRKSQBEKQLLRERERKRRQETDR 1294

RESULT 3
US-07-741-940-7
; Sequence 7, Application US/07741940
; Patent No. 5352775
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODIN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLIN, GROFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie & Beckett
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/741,940
; FILING DATE: 19920109
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.035574
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2842 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:

```

ORGANISM: Homo sapiens
IMMEDIATE SOURCE: CLONE: APC
US-07-741-940-7

Query Match 3.1%; Score 258.5; DB 1; Length 2842;
Best Local Similarity 17.9%; Pred.No. 4.4e-10;
Matches 304; Conservative 238; Mismatches 610; Indels 535; Gaps 69;

30 SHLTBEERKILAVMDROKKEEKEQSVLAKKEEKAQTFQWFFSGITELVNVNLOPOQ 89
1011 NEMDDNDGELDTPIVSLKYSDQ---LNSGRQSPSQNERWARPKHILB--DEIKSQE 1064
90 KQNEKEPOTKILHOQFEMITGEVKKMGESQ---QOEKQDADTCGICHYKFAQCGH 146
1065 ROSRQNS---TTPVYTESDTHKLFQPHFQOE---CVSPYRSRGANGSET 1111
147 NCSCYQTKFCARCGGCVSLRSKVMVNLCKQEIILTKSGAFYNSGNTLQOPQKV 206
1112 N-----RVGSHNGINQVNSQLC---QEDDYEDDKPTNYBERRYSEEEQHEE-- 1154
207 PRLGRNEBAQEKAKLHQPQFQCAPGLDLPVAVEKGRAGHLTRQDTIKNGSVKQIA 266
1155 -----KEPTNYSIKYNEKEKHVDQIDYSL-----KYA 1183
267 SMPDRKSPSVSRDQNRVRYQSEBEREDYQYVPSDGTMPSPSDVADRSGOBEPOFYE 326
1184 TQIPSSQKQSFQSKSSQSSQSKTE---HMSSEBENTSTPSSNAKQNLHPSSAQ 1236
327 EPGLNLYDNRGHRHSE---YIVDD-----EDVE 355
1237 SRSGQPKAATKCVSSINOETIQTTCVBDTPICFSCSSLSLSAEDEICGNTTQREAD 1296
356 SRDEYERQREFEYQARYSRDNLARYPVKPOPYEEQWRIHAFVSRKHERHSDVSLAN 415
1297 SANTLQIAIKIKICTRSAEDPVSEPAVSQHPRTKSSRLQGS-SLSESARHKAVERSS 1355
416 ABLEDSRISLLMDRP-----SRQSVSEERPAAMENQKVSMEKTRBAQ----- 459
1356 GAKSPKSCAQTPKSPPEHYVQETPLMFSRCTSVS---SLDSFESRSIASSVQSEPCSG 1411
460 -----GQSYQORTNSHPPTPRAPRPIPLDRPMDRRADSLRKHLLDPSSAVNKTREK 513
1412 MVSGIISPDLDPSQQTTPPSPRSKTPPP-----PPQTA--QTKRE- 1450
514 METMLANDSLSDQSESVRPPPPPHKSKGKGMQVSLSSSEEL-----ASTPEYTC 568
1451 ---VPKAPPAKTES-----GFKQAVNAVQVQVLPDADTLHFATESPDGESC 1501
569 D-----DVEL-----ESESVEKGDQ--KGRKKT--SEQG 595
1502 SSSLSALSLEPFIQDVELRIMPVQENDNGNETESQPKESMENQEKBAKTIDSEKD 1561
596 VLSDSN-----TSRQKRMVYGHSLBEDLEWSHQIKDQSGVDTCS 639
1562 LLSDSDDDDIIELECIISAMPTKSRKKAQPAQASKLPPPVARKPSQL-----PVYK 1615
640 TTLNEHSHSDKHPYVQSKQDRL--IGRILLN-----KRLKQSVPRDSCAMGL 690
1616 LUPQNLQPKQH-VSFTPGDMPRYCVGPTPIPFSTATSLDITSPSPNELAAGEV 1674
691 KVVGGKMTSGRLCAPIKVKKGLADTVGHLRPGDQVLEWNGRLLOCAATEEVYNTILE 750
1675 R--GG--AQSG-----EPKEDTIPTG--RSTDEA-----QGGKTSVITPILD 1713
751 SKPEPQVELV-----SRPIGD-----IPRIPDSTHAQLSSSSSFESQKMDRPSISVT 799
1714 DNKABEGDILACINSAMPKSKHPFRVKIMDQVQ-QASASSAPAKNQLDGGKKKPT 1772
800 SPMSPGMLRDVQFLSGQLSIKLFKVKHQHQLVILGAKDLPREDGRPNPVYKIYFL 859
1773 SPVKP-----IPQNTYRTRV-----RKNADSKNNLNAERVF 1804

860 PDRSDKMKRKYVKTLEPK-----WQOTFYSPVH-----RR 893
1805 SMDKSKKONLKONSKDFNDKLPNNBDRVGRSFAFDPHYHTTFFGYTCFSRNDSSL 1864
894 EFRERMLBITLWDQARVRESSEFLGELILETALLDDESPWVKLCTHDVSSLEPLPR-- 951
1865 DFDDDDVLSR-EKALEKAKENKESEAKVTSHTLSNQSANKQA--LAKQPINQGO 1921
952 PFTYLPBQOL-----HGESFTRRLQR-----SKRISDSRVEDYDCE----- 987
1922 PKPILQKSTPQSSKOIPDGAATDEKQALFALNTPVCFSHNSLSLSLSDIDQNNK 1981
988 -----DQGVVSDYRHNG-----RLQSSSTLSTVPEQVMSNVHC 1020
1982 ENEPKETEPPDSQEPKQAGVAPKSPHVEDTPVCFKSNLSLSLSDIDQNNK 2041
1021 SPGSGPHVD---VLGTRSNAPSAPPORNAVQCHGTRATGYNTVISM-D-RHRVMD 1076
2042 ISSAMPKKKYSRLKGDNEKISP-----RNM-QGILGEDLTLDKQIPDSEHGLSPD 2094
1077 HYSSDRDRDCEAAD-----RQ-----FYHRSSTE 1101
2095 SENFQWALQEGANSIVSSLHQAACLSRQASSDSLSLSKSGISLGSFHLTPDQ 2154
1102 QPILL-----KRTTSSRSRSPDTN----- 1122
2155 EKPFTSNKGPRLKPEKSTLTETTKIESSEKIGKGVKYSLTGVRNSSEISQWQ 2214
1123 -LWRSMPLMTGRSAPPSPALSRHPRGTV-----QTSPTSPTGVRGRGQLPOLPP 1174
2215 PLOANPISRGRTMIHLPGVNSSSSTSPVSKGPKLTPAKSPSEB---QTATSP 2270
1175 KOTLERSAMDIEERNOKMLNKYQVAGSDPLRBDYHSKYTRSG---WDPHRGAD----- 1226
2271 RGAKEPVKSELSPVARQT-----SQIGSS-----KAPSGSGRSTSPRPAQPLSR 2318
1227 TVSTSSSDSDVDSAVRTSASRFSSTSVMSVQSEPRGRKLSVFTSMQNRQMGVS 1286
2319 PIQSPGRNISPRNGISPPNKLSQLPRTSPSPSTASTKSGSGKWS-YTS-PCRQW--S 2373
1287 GKNLTKSTISGDMCSLEKNDQSDTAVGALGTSGKRRSSICAKWVAIVGLSKRSEA 1346
2374 QCNLTQGLSKNASSIPRSESASGLNMNGNANKK-----VELSRMSSTK 2422
1347 SOLSOTEGGKKL-----RSTVQSTETGLAVENNMWTKRQASRE-- 1388
2423 SSGESDRSERPVLRQSTFIKEAPSTLRKLESASFESLSPSRPASPTRQOQTPV 2482
1389 -----DGSNNYS 1397
2483 LSPSLPDMSLSTHSS 2497

RESULT 4
US-08-289-548A-7
Sequence 7, Application US/08289548A
Patent No. 5648212
GENERAL INFORMATION:
APPLICANT: ALBERTSEN, HANS
APPLICANT: ANAND, RAKESH
APPLICANT: CARLSON, MARY
APPLICANT: GRODEN, JOANNA
APPLICANT: HEDGE, PHILIP J.
APPLICANT: JOSLYN, GEOFF
APPLICANT: KINZLER, KENNETH
APPLICANT: MARKHAM, ALEXANDER F.
APPLICANT: NAKAMURA, YUSUKE
APPLICANT: THLIVERIS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
NUMBER OF SEQUENCES: 102
CORRESPONDENCE ADDRESS:
ADDRESSES: Banner & Allessretti, LTD

STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/289,548A
FILING DATE: 12-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107,46943
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2842 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: APC
US-08-289-548A-7

Query Match 3.1%; Score 258.5; DB 1; Length 2842;
Best Local Similarity 17.9%; Pred. No. 4,4e-10;
Matches 304; Conservative 239; Mismatches 618; Indels 535; Gaps 69;

QY 30 SHLEERKILAVMDQKKEKROSVLKIIEHKAQPTQWPFPGITLVNVLQPOQ 89
DB 1011 NHMDNDGELDTINSLKYSDEQ---LNSGRQSPQNERWARPKHIE--DEIKQSEQ 1064

QY 90 KQNEKEPQTLHQEQEMYTEQVKKHGESQ---QEQKGDAPTQGICHKYFADCGH 146
DB 1065 RQSRNQS-----TTPVYTESTDDKHLKQPHFGQOE-----CVSPYRSRGANGSET 1111

QY 147 NCVCQCTKPCARCGRVSLRSNKVMVNCRLKQOEILTKSGAWFYNSGNTLQOPDKV 206
DB 1112 N-----RVGSNHGINQVNSQLC-----QEDDYEDDKPTNYSERYSEEECHEE-- 1154

QY 207 PRLNEEAPQEKAKLHQEQFGAPGDLSPVPAVEKGRAHGLTROPTIKNGSVKHOIA 266
DB 1155 -----EERTPNYSIKYNEKRHVQDIDYSL-----KYA 1183

QY 267 SDMSDEKRPSPSRDQNRYYEQSEEREDYSQVPSDGTMPSPSDYADRSQREPOFYE 326
DB 1184 TDTPSQKQSFQSKSSGQSSKTE-----HMSSESTSTPSSNAKQNLHPSAQ 1236

QY 327 EPGLNVRDNRGRHRSKE-----YIYDD-----EDVE 355
DB 1237 SRGQPOKATCKVSSINQETIQTVCYEDTPICFSRCSSLSLSSAEDEIGCQNTQBEAD 1296

QY 356 SRDEYERQREHYEQARYRBDPNLARYPVKQPYEQMRTHAEVSRARHRRHSVSLAN 415
DB 1297 SANTLQIAIEKIKIGTSAEDPVSEVPAVSPHPTKSSRLQGS--SLSSSARHKAVEFSS 1355

QY 416 ALEDRIISLLMDRP-----SRQSVSEERRAAMENQSYGMERTBAQ----- 459
DB 1356 GAKSPKSGAQTPKSPPEHYVQETFLMFSRCTSVS-----SLDSPSRSIASSVQSEPCSG 1411

QY 460 -----QSSVPORTSNHSPPTPRSPITLDRPDMRADSLRKQHLDPSSAVKTKREK 513
DB 1412 MVSGIISFLDPSFGQTMPPSRSKTPPP-----PPQTA--QTKRE-- 1450

QY 514 METNLANDSLSDQSESVPFPPPHKSKKQKQKQVLSSESEKL-----ASTTFTTC 568
DB 1451 ---VPAKTAFTAKRES-----GPKQAAVNAVQVQVLPDADTLHPATSTTDPGSC 1501

QY 569 D-----DVEL-----ESESVEKGBQ--KQKKT--SBQ 595
DB 1502 SSSLNALSLEDFIQOVLELIMPVQENDNGNETSEQPKESNENQEKAKYIDSEKD 1561

QY 596 VLSDSN-----TRSERQKRMYYGSHLEEDLEMSHPQIDSGVDVTCSS 639
DB 1562 LLDDSDDDDIIEECIIISAMPTKSRKAKQAQASKLPPPVAKPSQL-----PVYK 1615

QY 640 TTNLEHSHSDKHPYTWQPSKODRL--IGRILLM-----XRLDGQVPPDSSGMLGL 690
DB 1616 LALPSQRLQPKH--VSFTPGDDMPRYCVGEPFINFSTATSLSLDTISFPFNLAGEGV 1674

QY 691 KVVGGKMTESGRLCAPITKVKKSLADTVGHLRPGDVLWNGRLQAGATFEVYVNIIE 750
DB 1675 R--QG--AQSG-----EPEKDTIPTEG--RSTDEA-----QCKTSVVTIPELD 1713

QY 751 SKPEPQVELV-----SRPID-----IPRIDSTHAQLESSSSSFESQKMDRPSISVT 799
DB 1714 DNKAERGDIABECINSAMPKGSKHPRVKIKMDVQ--QASASSAPKWKOLDKCKKPT 1772

QY 800 SPMSPGMLADVPOFLSGQLSIKMLFKVCHQLVITLIGADLPSEDDGPRNPYKLYFL 859
DB 1773 SPVKP-----IPQNTSYTRV-----RQADSKNNLNAERVF 1804

QY 860 PORSKDKERTKTKVKKTLSPK-----NNQTFYSPVH-----RR 893
DB 1805 SNNKSKKQNLKNSKDFDKLPNNEDRVGSAFDSPHYTFIEGTTCFSDNSSL 1864

QY 894 EFRERMLEITLMDQKRVRESEFLEILILETALLDDPHWYKLTQHDVSSLPR-- 951
DB 1865 DFDODDVLDR--EKAELKAKENKSEAKVTSTELTSQOSANKQA--IAQPTNRQ 1921

QY 952 PEPYLPFRQL-----HGESPTRLQR-----SKRISDSVSDYDCE-- 987
DB 1922 PKPILQKQSTFTPOSSNDIPDRGAATDEKLOFAIENTPVCFHNSLSLSLSDIQENNK 1981

QY 988 -----DGWGVVSDYRNG-----RDLQSSTLSVPRQVMSNHC 1020
DB 1982 ENEPIKETPPDSQGBSPKQASGVAPKFHVEDTTPCFERNSSLSLSDSDLLQEC 2041

QY 1021 SPGSPHRYD---VIGRTSWSFSPAPPQORVQERGRATRYNTYISMD--RHYVMD 1076
DB 2042 ISSAMPKKKPKSLKGDNEKHSB-----RNM--GJLGEDLTLDLKDQIPDSEKGLSPD 2094

QY 1077 HYSDDRDRCBAAD-----RQ-----PTHRGRSTE 1101
DB 2095 SENFDWKAIQEGANSIVSSLHQAAAACLSRQASDSDSILSLKSGISLSPHLPDQE 2154

QY 1102 QRELL-----ERTTTSRSESRPDTN----- 1122
DB 2155 EKPFTNGKPRILKPEKSTLETKLESKSGIKGKQKVTKSLITKVRNSISGOMKQ 2214

QY 1123 -LMSNPSLMTGSAAPPSPALSHRPHQSV-----QTSPESTPTQYGRGRQLPOLPP 1174
DB 2215 PLQANPISIRSGRTMTHIPGVNRSSSTSPSVKKGPPKLPKTPASKSPSEB---QTATTSP 2270

QY 1175 KGTLESAMDIERENOMKLNKYQVAGSDPRLQDYHSTYRSG---NDHPRGAD----- 1226
DB 2271 RGAKPSVKSELSVARQT-----SQIGSS-----KAPSRSGSRDSTPSRPAQPLSR 2318

QY 1227 TVSTKSSDSDVSDVSAVSTSSASRFSSTYSMSVQSERPRGNEKIVPTSKMNRQMGVS 1286
DB 2319 PIQSPGRNISLSPNGIGISPPNKLSQLPRTSSPSTASTKSGSGKMS--YTS--PGROM--S 2373

QY 1287 GKMLTKSTISGDMCSLEKNDGQSQSDTAVGALGTSGKGRSSITGAKWVAIVGLSRSSA 1346
DB 2374 QONLTKQTLGSKNASSIPRSESASGLNQNNNGANKK-----VELSRSSSTK 2422

QY 1347 SOLSQTEGGKKL-----RSTVORSTETGLAVEMRWMTROAREST----- 1388
 Db 2423 SSGSESDREPRVLRQSTFIKEAPSTLRRKLEESASFESLSPSSRPASPTRSQATPV 2482
 QY 1389 -----DGSMSYSS 1397
 Db 2483 LSPSLPMSLTHSS 2497

RESULT 5
 US-08-452-654-7
 ; Sequence 7, Application US/08452654
 ; Patent No. 5691454
 ; GENERAL INFORMATION:
 ; APPLICANT: ALBERTSEN, HANS
 ; APPLICANT: ANAND, RAKESH
 ; APPLICANT: CARLSON, MARY
 ; APPLICANT: GRODEN, JOANNA
 ; APPLICANT: HEDGE, PHILLIP J.
 ; APPLICANT: JOSLYN, GLOFF
 ; APPLICANT: KINZLER, KENNETH
 ; APPLICANT: MARKHAM, ALEXANDER F.
 ; APPLICANT: NAKAMURA, YUSUKE
 ; APPLICANT: THLIVERIS, ANDREW
 ; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
 ; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
 ; NUMBER OF SEQUENCES: 94
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSES: Banner, Birch, McKie & Beckett
 ; STREET: 1001 G Street, NW
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20001-4598
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/452,654
 ; FILING DATE: 25-MAY-1995
 ; CLASSIFICATION: 536
 ; PRIORITY APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/741,940
 ; FILING DATE: 08-AUG-1991
 ; TELEPHONE: 202-508-9100
 ; TELEFAX: 202-508-9299
 ; INFORMATION FOR SEQ ID NO: 7:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2842 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; ORIGINAL SOURCE:
 ; ORGANISM: Homo sapiens
 ; IMMEDIATE SOURCE:
 ; CLONE: APC
 ; US-08-452-654-7

Query Match 3.1%; Score 258.5; DB 1; Length 2842;
 Best Local Similarity 17.9%; Pred. No. 4.4e-10;
 Matches 304; Conservative 238; Mismatches 618; Indels 535; Gaps 69;
 QY 30 SHLTEREKILLAVMDQKKEEKESQSVLAKKEHKAQPTQWFFFSGITELVNVVLQPPQ 89
 Db 1011 NHMDNDGELDTPIYSLKYSDEQ-----LNSGRQSPQNERWARPKHIE--DEIKQSEQ 1064

QY 90 KQPNKEPOTKLHQFQFMYKQVKKMGHESQ-----QSQKGDAPYCGTCHKTKPADQCGH 146
 Db 1065 QSRNQS-----TTYPTVYTESDDKHLKQPHFQOE-----CVSPYASRGANGSET 1111
 QY 147 NCSTYQOTKFCARCGRVSLRNVWVNCNLCKKQEIILTKSGAMPYNSGNTLQPDQKV 206
 Db 1112 N-----RVGNNHGINQVQSGLC-----QDDYEDDKPTWYSRYSEERQHE-- 1154
 QY 207 PRGLENEBAPOEKKAKLHEQPOFOGAPGDLSPVAVEKRAHGLTRQDTIKNGSGVKHOIA 266
 Db 1155 -----EERPTWYSIKYNEEKHVDQPIDYSL-----KVA 1183
 QY 267 SDMPDRKSPSVSRDQNRRYEQSEEREDYQVPSDQTMFRSPDYADRSQREPOPYE 326
 Db 1184 TDIPSSQKQSFPSKSSGQSKTE-----HMSSSESTSTPSSNAKRONLHPSSAQ 1236
 QY 327 EPGHLNYRDSNRGRHSHKE-----YIVDD-----EDVE 355
 Db 1237 SRSGQPOKAACTKVSSINQETIQTVCVEDTFCFRCSSLSLSAEDEICGNTQTQAD 1296
 QY 356 SRDEYERQREERYQARYRSDFNLARYPVKQPYEEOIRIHAEVSRARHERHSDVSLAN 415
 Db 1297 SANTLQIARIKEIKIGTRSAEDPVSEPAVSQHPTKSSRLQGS--SLSSSAHKAVERSS 1355
 QY 416 AELEDSRISLLMRDP-----SRQSVSERRAAMENQBSYSHVERTBAQ----- 459
 Db 1356 GAKSPKSGAQTPKSPPEHYVQETPLMFSRCTSVS-----SLDSPSRSIASSVQSEPCSG 1411
 QY 460 -----GQSSYQRTSNHSPPTFRSPPLDRPDMRADSLAKQHLDPSSAVRATKREK 513
 Db 1412 MVSGIISPDLDPSGQTMPPSRKTPPP-----PPQTA--QTKRE- 1450
 QY 514 METMLRNDLSLSDQSESVRPPPRPHKSKGKGMQVSLSSSEEL-----ASTPYTSC 568
 Db 1451 ---VPKNKAPTAKRES-----GPKQAAVAAVQKVQVLPDADTLHPATESTPDGFSK 1501
 QY 569 D-----DYEL-----ESESVEKGDQO--KGRKKT--SEOG 595
 Db 1502 SSSLGALSLEDPFIQKQVELRIMPVQENDNGNETESEOQKESNENQEKAEKTIIDSEKD 1561
 QY 596 VLSDSN-----TRSEROKRMVYGHSLSEELWSEPOIKDGVDTCSG 639
 Db 1562 LDDDDDDDIRILEECIIISAMTKSSRKAQPAQATASKLPPPVAKPQSL-----PVTK 1615
 QY 640 TTLNEESHSDKHPVTWQPSKQDGL--IGRILLN-----KRLKQGVPRDSGAMLSL 690
 Db 1616 LLPSQNLQPKH-VSFTPGDDMPRYCVGTPINFSTATSLDLTIESPPNELAAGECV 1674
 QY 691 KVVGKMTESGRLCAPITKVKKGLADTVGHLRPGDEVLEWNGRLLOQATFEEVYNIILR 750
 Db 1675 R--GG--AQSG-----EFKRDITPTG--RSTDEA-----QGGKTSVVTIPELD 1713
 QY 751 SKPEQOVELVW-----SRPIGD-----LPRIPDSTHACLESSSSSFESEKMDRPS18VT 799
 Db 1714 DNKAEGDILAEICINSAMPKSHKPRVKLMQVQ--QASASSAPKWKQDGGKKKKT 1772
 QY 800 SPMSFQMLRDVQFLSGQLSIKLFDFKVGHLIVTILGAKOLPREDGDRPNPYVKIYFL 859
 Db 1773 SPVKP-----IPQNTYRTRV-----KKAADSKNLAERYP 1804
 QY 860 PDRSDKNKRTTKVKTLEPK-----WQYTIYSPVH-----RR 893
 Db 1805 SDNKSCKQNLKNSKDFNDKLPNNEDRVGRSFAFDSPEHYTF1EGTPYCFGRNDSLSSL 1864
 QY 894 EFERMLEITLMDQARVREESEFLGEILILELTALDDPHWYKLOTHEDVSLPLPR-- 951
 Db 1865 DDDDDDDVDSR-EKAEIRKAKENKSEAKVTSHTLTSTNQSQANKTQA--IAQPINRQO 1921
 QY 952 PSPYLPRRQL-----HGSEPTRLQR-----SKRISDSVSDYDCE-- 987
 Db 1922 PRFLLQKQSTFFQSSKIDPRGAATDEKLNPAIENTPVCFSHNSLSLSLSDIDQNNNK 1981

988 -----DCGVVSDYRHHG-----RDQSSTLSPVPEQVMSNHC 1020
1982 ENPEIKETEPDQGGSPKQAGYAFKSHVEDTVPVCFSRNSSLSDSDLLQBC 2041
1021 SPSSGSHRYD---VIGRTSWSPSAPPQPNVQCHGTRATYNTISMD-RHRVMD 1076
2042 ISSAMPKKKPSRLKGNKHSPP-----RNM-GGILGEDLTLDKDIQEPDSEHGLSPD 2094
1077 HYSDDRRDCEAD-----RQ-----PYHRSRST 1101
2095 SENFDWKAIQEGANSIVSSLHQAACLSRQASSDSDSLSLKSGISLSPHLPDQ 2154
1102 QRELL-----BRTTTRSSSRPTN----- 1122
2155 EKFTNKGPRILKPKGKSTLETKLESKIGKGGKVKYSLITGVRSNLSISOMQ 2214
1123 -LMSMPSLMTGSAAPPSPALSHPRVTSV-----QTSPTSTGTGRRGRQLPOLP 1174
2215 PLQANMPSISGRTHIHPVRNSSLSTSPVSKGPKLTPASKSPSG-----QTATTSP 2270
1175 KGTLSAMDIERNOMKLNKYQVAGSDPRLQDYHVKYRSG---MDPHRGAD----- 1226
2271 RGAKPSVKSELSVARQT-----SQIGSS-----KAPSRSGRSDSTSPRPAQPLSR 2318
1227 TVSTKSSDSDVDSAVRSTSSARFSSSTSYMSVQSERPRGNEKISVFTSKMOROMGVS 1286
2319 PISPGNRNLSIPGRNLSPPNKLSQLPRTSSPTASTKSGSGKMS-YTS--PGRQM--S 2373
1287 GKNLTKSTISGDMCSLEKNDGQSDTAVGALGTSGKRRSSIGAKWVAIVGLSRKSA 1346
2374 QNLTKQTGLSKNASSIPRESASKGLNQNNNGANK-----VELSRMSSTK 2422
1347 SLSQTEGGKKL-----RSTVORSTETGLAVRMNMTQASRST----- 1388
2423 SSGSDRSEPRVLRQSTFIKEAPSTLARKLESASPESLSPSRPASPTRSQATPV 2482
1389 -----DGSNNYS 1397
2483 LSPSLFDMLSLTHSS 2497

RESULT 6

US-08-449-731-7

Sequence 7, Application US/08449731

Patent No. 6413727

GENERAL INFORMATION:

APPLICANT: ALBERTSEN, HANS

ANAND, RAKESH

CARLSON, MARY

GRODEN, JOANNA

HEDGE, PHILIP J.

JOSLYN, GREGG

KINZLER, KENNETH

MARKHAM, ALEXANDER F.

NAKAMURA, YUSUKE

THLIVERIS, ANDREW

TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC

GENE IN COLORECTAL CANCER IN HUMANS

NUMBER OF SEQUENCES: 102

CORRESPONDENCE ADDRESS:

ADDRESSEE: Banner & Allegretti, LTD

STREET: 1001 G Street, NW

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20001-4598

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/449, 731

FILING DATE: 25-May-1995
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/289,548
FILING DATE: 12-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107,46943
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2842 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: APC
SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-08-449-731-7

Query Match 3.1%; Score 258.5; DB 4; Length 2842;
Best Local Similarity 17.9%; Pred. No. 4.4e-10;
Matches 304; Conservative 238; Mismatches 618; Indels 535; Gaps 69;
QY 30 SHLTERRKIIIAVMDRQKKEEKEKQSLVKKKEHKAQTPWPPSGITELVNVLPQQ 89
DB 1011 NHMDNDGELDTPIINSLAYSEQ-----LNSGRQSPQWERWAPKHTE--DEIKQSQ 1064
QY 90 KPNKEKPTQLHQFQFMYKEQVKMGESQ---QORQKQDAPTGGICHTKTFADGGH 146
DB 1065 RQSRNQS-----TTPVYTESTDDKHLKQPHFQQG-----CVSPSRGANGSET 1111
QY 147 NCYQCTKFCARCGGRVSLRNKVMVNCMLCKQKQKILTKSGAMFNKSGNTLQDPQKV 206
DB 1112 N-----RVGNHGINQVNSQLC-----QEDDYEDDPTWYSEYSEBOHER-- 1154
QY 207 PRGLNNEAPQKKAKLHBPQFQCAPGDLSPVAVKGRAGLHTRQDTIKNGSGVKHQIA 266
DB 1155 -----EERTVNSIKYNEKRVDPQIDYSL-----KYA 1183
QY 267 SDMPDRKESPSVRDQNEREYQSEEREDYSYVPSDGTMPSPSPDYADRSQRERQFYE 326
DB 1184 TDIPSQKQSPFSKSSGQSKTB-----HMSSESENTSTPSSNAKRONQLHPSSAQ 1236
QY 327 RFGHLNTRDSNRGRHRSKE---YIVDD-----EDVE 355
DB 1237 SRSGQPQKAATCKVSSINQETIQTVCVETPTCPSCSLSLSSAEDIGCMTQTOEAD 1296
QY 356 SRDEYERORREYQARYSDPNLARYPKVPQFYEQWRIHAEVSRARHERHSDVSLAN 415
DB 1297 SANTLQIARIKEKIGTRSAEDPVSPVAVSQHPRTKSSRLQGS-SLSSESARHKAVERSS 1355
QY 416 AELEDSRISLLMRDP-----SRQSVERRAAMENQSYSMERTBAQ----- 459
DB 1356 GAKSPKSGAQTPKSPFPEHYVQSTPLAFSRCTSVS-----SLDSFESRSIASVQSEPCSG 1411
QY 460 -----GQSSYPORTSNHSPPTPRSPPLDRPDWRADSLKQHLHLPSSAVRKTREK 513
DB 1412 MYSGIISPSDLPSQGTMPFBSKTPPP-----PPTTA--QTKRE- 1450
QY 514 METMLRNDLSLSDQSESVRPPPPPHKSKKGKGMKQVSLSSSEHEL-----ASTTETTSC 568
DB 1451 ----VPKNKAPTAEKRES-----GPKQAANAQVRQVLPDADTLHFAFATSTFDGFS 1501
QY 569 D-----DVEL-----ESSVSVEKGDQ--KGRKKT--SEGG 595
DB 1502 SSSLGALSILDEPFIQKDVLELIMPPVQENDNGNETESQPKESNENQKAEKTTIDSEK 1561

596 VLSGDK-----TRSRQKRMYYGHSLEEDLEWSBPQKDSGVDTCS 639
 1562 LLDSDDDDIIEECIIISAMPTKSSKAKAQAATASKLPPPVAKPSQL-----PVYK 1615
 640 TTLNEZSHSDKHPTWQSKDRL--IGRILLN-----KLLKDSVPRDSGMLGL 690
 1616 LLPSQNLQPKH-VSFTFGDDMPRYCVGTFINFSTATSLDITIESPPNELAAGGV 1674
 691 KVVGGKWTESGRICAFITTKVKGSLADTVGHLPKPGDEVLEWNGRLQAGATFEVYNIL 750
 1675 R--GG--AAGG-----EFKEDTITPEG--RSTDEA-----QGGTSSVTIPELD 1713
 751 SKPEQVELVY-----SRPTGD-----IPRPDTHAQLESSSSFESQKMDRPSISVT 799
 1714 DNKAEBGDIILAEICINSAMPKGSHPKPVKIMQVQ-QASASSAPNKNQLDGKKKPT 1772
 800 SPMSFGMLRDVQFLSGQSLKLPKDKVGHQILVITLGAOLPSRDGPRNPVYKIYL 859
 1773 SPVKP-----IPQNTYRTKV-----RKVADSKNMLNARVF 1804
 860 PORSKXKRTKVKITLSPK-----WNOTFIYSPVH-----RR 893
 1805 SDKDSKKQNLKNSKDFDKLPNNEDVRGSAFDPSPHYTIEGTPYCFERNDSLSSL 1864
 894 EFRERMLEITLMDQARVREESBFLGELILETALLDDEPHWYKLOTHDVSSLPRLP-- 951
 1865 DFDDDDVLSR-EKALERKAKENKSEAKVTSHTLTSNQSANKTQA--IAKQPINRQ 1921
 952 PSYLPRLRQ-----HGESPTRELOR-----SKRISDSVSDYDCE----- 987
 1922 PKPILQKSTFPQSSKIDPRGAATDEKLNQFAIENTPVCFSHNSLSLSIDSDENNK 1981
 988 -----DGVGVVDYRNG-----RDLQSTLSVPEQVMSNEC 1020
 1982 ENEPKETEPPDQGBRKSQASGYAPKSPHVEDTPVCFSRNSLSLSIDSDILLQEC 2041
 1021 SPGSGPHRVD---VIGRTRSWSPSAPPQPNVRQGRGTRATGHYNTISMD--RHRVMD 1076
 2042 ISSAMPKPKKPSRLKGNKHSK-----RNM--GGILGEDLTLDKDIQRPDSHGLSPD 2094
 1077 HYSDDRRDCEAD-----RQ-----PHERSSTE 1101
 2095 SENFDWKAIOGANSIVSSILHQAACLSQASSDSLSLSKSGISLSPHLPDQ 2154
 1102 QRELL-----ERTTRGRKSERPDN----- 1122
 2155 EKFTSNKGRPIIKPGEKSTLETKTESKGIKGGKVKYSLITGKVRNSLSISQMKQ 2214
 1123 -LMRSMPSLMTGRSAPPSPALSRHPRTGSV-----QTSPTSTGTGRRGRLPOLPP 1174
 2215 PLQANWPSISGRRTIHIPVRNSSSTSVPKKGPPKTPASKSPSEG---QTATTSP 2270
 1175 KGTLESAMDIERNRQMLNKYQVAGSDPRLEODYHVKYRSG---WDPRGAD----- 1226
 2271 RGAKPSVKSLSVPARQT-----SQTGGSS-----KAPSRSGSRDSTPSRPAQPLSR 2318
 1227 TVSTKSDSDVDSVAVSRSTSSASRPSSTVMSVQSERPRGNKISVFTSKMQRQMGVS 1286
 2319 PIQSPGENSISPRNGISPPNKLSQLPRTSPSTASTKSSGSKMS-YTS--PCRQ--S 2373
 1287 GKNLTKSTSIGDMCSLEKNGSQSDTAVGALGTSGKGRSSISGAKMVAIVGLSRKRS 1346
 2374 QONLTQGLSKNASSIPRSASAKGLNOMNGANGNK-----VELSRMSSTK 2422
 1347 SOLSQTGGGKKL-----RSTVORSTETGLAVENRMTQASREST----- 1388
 2423 SSGESDRSRPVLVQRSTFKAPSTPLRKLKESASFESLSPSRPASPTRSQATPV 2482
 1389 -----DGSNVSYS 1397
 2483 LSPSLPDMSLSTHSS 2497

RESULT 7
 US-08-452-655B-2
 : Sequence 2, Application US/08452655B
 : Patent No. 5783666
 : GENERAL INFORMATION:
 : APPLICANT: ALBERTSEN, HANS
 : APPLICANT: ANAND, RAKESH
 : APPLICANT: CARLSON, MARY
 : APPLICANT: GRODIN, JOANNA
 : APPLICANT: HEDGE, PHILIP J.
 : APPLICANT: JOSLYN, GEOFF
 : APPLICANT: KINZLER, KENNETH
 : APPLICANT: MARKHAM, ALEXANDER F.
 : APPLICANT: NAKAMURA, YUSUKE
 : APPLICANT: TILVERIS, ANDREW
 : TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
 : TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
 : NUMBER OF SEQUENCES: 102
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Banner & Witcoff, Ltd.
 : STREET: 1001 G Street, NW
 : CITY: Washington
 : STATE: D.C.
 : COUNTRY: USA
 : ZIP: 20001-4598
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patent In Release #1.0, Version #1.25
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/452,655B
 : FILING DATE: 25-MAY-1995
 : CLASSIFICATION: 530
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 08/289,548
 : FILING DATE: 12-AUG-1994
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 07/741,940
 : FILING DATE: 08-AUG-1991
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Kagan, Sarah A.
 : REGISTRATION NUMBER: 32,141
 : REFERENCE/DOCKET NUMBER: 1107.49964
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: 202-508-9100
 : TELEFAX: 202-508-9299
 : INFORMATION FOR SEQ ID NO: 2:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 2843 amino acids
 : TYPE: amino acid
 : TOPOLOGY: linear
 : MOLECULE TYPE: protein
 : US-08-452-655B-2
 : Query Match 3.1%; Score 258.5; DB 1; Length 2843;
 : Best Local Similarity 17.9%; Pred. No. 4.4e-10;
 : Matches 304; Conservative 238; Mismatches 618; Indels 535; Gaps 69;
 : QY 30 SHLTEREKILIAVMDRQKKEKESQVLKKEHKAQTOFWPPFGITELVNVVLPQQ 89
 : Db 1012 NHMDNDGELDTPIYSLKYSDEQ-----LNSGRQSPSONERWARPKHTE--DEINQSEQ 1065
 : QY 90 KQPNKEPQTLHQPEYKQVKKMGESQ---QKQKGDAPTCGICHKTKPADGCGH 146
 : Db 1066 ROSRMSQ-----TTPVTTESTDDKHLKPQHPFGQEE-----CVSPYRSGANGSET 1112
 : QY 147 NCSYCQTKFCARCGRVSLSRKNVVCNLCRQQLTKSGAWFYNSGNTLQQPDQKV 206
 : Db 1113 N-----RVGSNHGINQVNSQLC-----QEDDYEDDKPTWYSEYSEEEQHEE-- 1155
 : QY 207 PGLRNERAPQEKAKLHBPQFQAGDLSVPAVEKGRAHGLTROTDTIKNGSVKXQIA 266

Db 1156 -----EERTNYSIKNEKRVWDQIDYSL-----KYA 1184
QY 267 SDMPDRKSPSVSRDQNRREYEREDYEQVPSDGTMPRSPDYADRRSQREPOFYR 326
Db 1185 TDIPSSQKQSPFSKSSGSSKTE-----HMSSESTSTPSENAKQNLHPSSAQ 1237
QY 327 EPGLNTRDNRGRHRSKE-----YIVDD-----EDVE 355
Db 1238 SRSGQPOKATCKVSSINQETIQTYCVEDTPICFSCSSLSLSSAEDEIGCQYTOEAD 1297
QY 356 SRDEYERQREEEVQARYRSDPMLARYPKPQVPEEQRIHAYVSRARHRRERESDVLAN 415
Db 1298 SANTLQIAEIKKIGKTSASDVPSEVPAVSGHPRTKSSELOGS-SLSSERARKAVFSS 1356
QY 416 AELEDRSISLLMRDP-----SRQSVSEKRAAMENORSYMERKTEAO-----459
Db 1357 GAKSPKSGAQTPKSPPEHYQETPLMFSRCTVS-----SLDSFERSIASVQSEPCSG 1412
QY 460 -----QSSYPORTSNHSPPTPRRSPIPLDRPDMRRADSLRKQHLDPSSAVKTKREK 513
Db 1413 MVSGLIISPSDLPSPGQTPMPSRSKTPPP-----PPQTA--QTKRE- 1451
QY 514 METMLRNDLSLSSQSESVRPPRPKSKKGGKMRQVLSLSSSEEL-----ASTPYTSC 568
Db 1452 ---VPKNKAPTAKRES-----GPKQAAVNAAVQVRVQVLPDADTLHAFATESTPDGFSC 1502
QY 569 D-----DVEL-----ESESVESEKGDQO-KGKRTK---SEOG 595
Db 1503 SSSLSALSILDEPTIQDVELRIMPVQVNDNGNETSEQKPEKNEQKEAKETIISKED 1562
QY 596 VLSDSN-----TRSERQKRMVYGGHSLSEEDLEWSEPOKIDSGVDTCSS 639
Db 1563 LLDDSDDDDEIIEBECIISAMPTKSKKAKPACTASKLPPVARKPSQL-----PVYK 1616
QY 640 TTLNEESHESDKHPTWQPSKGDRL---IGRIILN-----KRLKDGSVPRSGAMJGL 690
Db 1617 LLPSONRLOPQKH-VSFTPGDDMPRVYCVGTRINFSTATSLDATESPPNELAAGEV 1675
QY 691 KVVGGKMTESGRICAFITTKVKKSLADTVCHLRPGDVLWNGRLQCATFEVYVNIILR 750
Db 1676 R--GG--AQSG-----EPEKDDTIPTEG--RSTDEA-----QGGKTSSTVTIPELD 1714
QY 751 SKPEQVELVW-----SRPIGD-----IPRIPISTHAQLESSESPESQKMDRPSISVT 799
Db 1715 DNKAEBGDILAEICINGAMPKGGKHPKPVCKIMDQVQ-QASASSAPNKNQLDGKKKPT 1773
QY 800 SPMSPGMLRDVPFLSGQLSIKLPDKVGHQILVITLIGAKDLPSREDGRPNPYKIYFL 859
Db 1774 SPVKP-----IPQNTYRTRV-----RKNADSNNLNAERVF 1805
QY 860 PDRSDKMKRTKTQVKKTLBPK-----MNQTFIYSPVH-----RR 893
Db 1806 SDNKSKKQNLKNSKDFNDKPNNEEDRVGSPAFSPHYTIFBTGTCFSDNSLSL 1865
QY 894 EFRERMLEITLMDQARVRESEFLGEILTELTALLDDPHWYKLTQTHDVSSLPFR-- 951
Db 1866 DFDDDDVDLSR-EKALRKAKENKSEAKVTSHTLTSNQQSANKQA--IAQPINRGQ 1922
QY 952 PSYLPFRQI-----HGSPTRLOR-----SKRISDSSEVSDYCB-----987
Db 1923 PKPILQKSTFPQSKDIPDRGAATDEKLQFAIENTPVCFSHNSLSLSLSDIDQENNNK 1982
QY 988 -----DGVGVVSDYVRHG-----RDLQSTLAVPEQVMSSNHC 1020
Db 1983 ENEPKETEPPDQGGSPKQASGYAPKSHFVEDTVPVCFSRNLSLSLSDSDLLQEC 2042
QY 1021 SPSSGSHRVD---VIGRTRSWSPSPAPPOQNVQHRGTRATGYNTYITGRMD-RHRVMD 1076
Db 2043 ISSAMPKCKKPSRLKGDNEKHS-----RNM-GGILGEDLTLDKDIQRPDSEHGLSPD 2095
QY 1077 HYSDDRDRCDAAD-----RQ-----PYHRSRST 1101

Db 2096 SENFOWKATORGANSIVSSILHQAACCLSRQSSDSDSILSKGSLGSPPHLTPDOE 2155
QY 1102 QRPILL-----ERTTTSRSESRPDTN-----1122
Db 2156 EKPFTSNKGPRLKPOEKSTLETKKLESKSGIKGOKKYKSLITGKVRNSISQOMQ 2215
QY 1123 -LWRSWPSLWTRGSAAPPSPALSRSHPTGSV-----QTSFSTPTGTQRRGRQLPPLP 1174
Db 2216 PLQANWPSISRGKRTMTHIPOVRNSSSTSPVSKGPPKLPKTPASKSPSEK---QTATTSP 2271
QY 1175 KGTLESAMDIEERNOMKLYKQVAGSDPRLQDYHKKYRSG---WDFRGAD-----1226
Db 2272 RGAKPSVKSLSFVARQT-----SOIGSS-----KAPSSGSRDSTPSRPAQOPLSR 2319
QY 1227 TVSTKSDSDVDVSAVSRSTSGASRPSSTSYMSVQSERPRGKIKSVFTSQMCRMGVS 1286
Db 2320 PIOSPGRNISIPOENGISFPKLSQLPRTSSPSTASTKSGSGKMS-ITS--PGRM--S 2374
QY 1287 GKMLTKSTISGDMCSLEKNGQSDTAVGALGTGCKKRRSISGAKOVAIVGLSRKRSRA 1346
Db 2375 QKMLTKQTLGKLNASSIPRBSASKGLNQMNNGANKK-----VELSRMSSTK 2423
QY 1347 SOLSQTEGGGKKL-----RSTVORSTETGLAVEMNMTRQASRST-----1388
Db 2424 SGSESDRSERFVLRQSTFIKEAPSTLRLKLESASFESLSPSRPASPTRSQAQTPV 2483
QY 1389 -----DGSMSYSS 1397
Db 2484 LSPSLPMSLSTHSS 2498

RESULT 8

US-08-452-655B-7
; Sequence 7: Application US/08452655B
; Patent No. 5783666
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GREGG
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER P.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 102
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Banner & Witcoff, Ltd.
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452,655B
; FILING DATE: 25-MAY-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/289,548
; FILING DATE: 12-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/741,940
; FILING DATE: 08-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.

APPLICANT: HEDGE, PHILIP J.
 APPLICANT: JOSLYN, GREGG
 APPLICANT: KINZLER, KENNETH
 APPLICANT: MARKHAM, ALEXANDER F.
 APPLICANT: NAKAMURA, YUSUKE
 APPLICANT: THLIVERIS, ANDREW
 TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
 TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
 NUMBER OF SEQUENCES: 102
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Banner & Witcoff, Ltd.
 STREET: 1001 G Street, NW
 CITY: Washington
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20001-4598
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/450,582
 FILING DATE: 12-AUG-1994
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/452,655
 FILING DATE: 25-MAY-1995
 APPLICATION NUMBER: US 08/289,548
 FILING DATE: 12-AUG-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/741,940
 FILING DATE: 08-AUG-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: Kagan, Sarah A.
 REGISTRATION NUMBER: 32,141
 REFERENCE/DOCKET NUMBER: 1107.49964
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202-508-9100
 TELEFAX: 202-508-9299
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2843 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-450-582-2

Query Match 3.1%; Score 258.5; DB 3; Length 2843;
 Best Local Similarity 17.9%; Pred. No. 4.4e-10;
 Matches 304; Conservative 238; Mismatches 618; Indels 535; Gaps 69;
 QY 30 SHLTERRKILLAVMDQKKEEKEQSVLAKIEHKAQPTQWPPFSGITELVNNVLPQOQ 89
 DB 1012 NHMDNDGELDTPIYSLKYSDEQ-----LNSGRQSPQNERNAPKHIIE--DEIKQSEQ 1065
 QY 90 KQNEKEPQTKLHQFQMYKEQVKMGKERSQ---QQEQKGDAPTCGICHTKTFADCGCH 146
 DB 1066 RQSRNQS-----TTYPVYESTDHLKLPQHPQOE-----CVSPYRSRANGSET 1112
 QY 147 NCSCYQTKFCARCGRVSLSRKNVKNVNCNCRKQOEILLTKSGAWFYNSGNTLQQPQKV 206
 DB 1113 N-----RVGNSHGINQNVQSILC-----QEDDYEDDKPTNYSEYSEBEQHEE-- 1155
 QY 207 PRGLENEEAQEKAKLHEQPFQAGPCDLSPVAVEKRAHGLTRQDTIKNGSGVHQHIA 266
 DB 1156 -----EERPTNYSIKNEKRHRVDQPIDYSL-----KYA 1184
 QY 267 SDMPDRKRGPSVSRDQNRREYQSEEREDYSQYVPSDGTWPRSPDYADRRSREPOFYE 326
 DB 1185 TDIPSSQKQSFSSKSSGSSKTE-----HMSSESTSTPSSNAKQNLHPSSAQ 1237
 QY 327 EPGLNYRDSNRGRHRSKE-----YIVDD-----EDVE 355

DB 1238 SRSGQPKAATCKVSSINQSTIQTVCVEDTPTCFSCSLSLSAEDIGCNQTTQEDAD 1297
 QY 356 SRDEYERQREERYQARYSDPWLARYPVKPYQPYEOMRIHARVSRARHERSDVSLAN 415
 DB 1298 SANTIQAIEIKELIGTRSAEDPVSEVPVAVSQRPTKSSRLOQS-SUSSSARHKAVERFS 1356
 QY 416 AELEDSRI8ILRMDRP-----SRQSVSERRAAMENQRYSMERTEAQ-----459
 DB 1357 GAKSPKSGAQTPKSPPEHVQSTPLWFSEKTSVS-----SLDSFESRSTASSVQSEPCSG 1412
 QY 460 -----GQSSYPORTSNHSPPTPRSPPIPLDRPMRADSLRKHLDLPSSAVKTKREK 513
 DB 1413 MYSGIISPDLPDPOQTWPPSRSKTTPP-----PQTA--QTKRE- 1451
 QY 514 METMLRNDLSLSDSSESVRPPPRPHKSKKGGMROVLSLSSSEKEL-----ASTPEVTS 568
 DB 1452 ----VPKKAFTAKRES-----GPKQANAAVQVQVLPDADTLLHPTATSTPPGFC 1502
 QY 569 D-----DVEL-----ESESVEKGDQ--KQKRT--SEOG 595
 DB 1503 SSSLSALSLEPFIQKDVLRIMPPVQVNDNGNETSEKQPKSENQEKAKTIDSEKD 1562
 QY 596 VLSDSH-----TRSERQKRMVYGHSLSEEDLEWSEPKQKSGVDTCSG 639
 DB 1563 LLDDSDDDDIIEILECIIISAMPTKSSRKAKPAQTASKLPPPVARKPSOL-----PVYK 1616
 QY 640 TTLNEHSHSDKHPTWQPSKGDRL--IGRIILN-----KRLKDGSPRDSGAWLGL 690
 DB 1617 LLPSQRLQPOKL-VSFTPODDMPRVYCVGPTINFSTATSLDITREPPNELAAGEV 1675
 QY 691 KVVGGMTESGRICAFITVKKGLADTVGHLRPGDEVLEWNGRLQAGATFEVYNIILE 750
 DB 1676 R--GG--AQSG-----EFKEDTITTEG--RSTDEA-----QGGKTSSTVITPELD 1714
 QY 751 SKPEQVELV-----SRPID-----IPRIDSTHAQLESSESSFFESQKMDRPSISVT 799
 DB 1715 DNKAEGDILABCINSAMPKSGSHKPFVKIKMDVQV-QASASSAPFNKQLDKKKKPT 1773
 QY 800 SPMSQGLRDVQFLSGQLSILKLPKVGHLVITLIGAKLPSEDRGPRPNVVKIYFL 859
 DB 1774 SPVKP-----IPONTYRTRV-----RKAQSKNNLNARVP 1805
 QY 860 PDRSDKNKRTKTKVKTLEPK-----NNQTFYSPVH-----RR 893
 DB 1806 SDNKSQKQNLKNSKDFNDKLPNNEDVRGSPAFDPSPHYTFIEGTPYCFSRNDSLSL 1865
 QY 894 EFRERMLEITLMDQARYRESEFZGLIELLETTALDDPHYKLOTHDVSLPLPR-- 951
 DB 1866 DFDDDDVDLSR-EKAEILKAKENKESEAKVTSHTLTSTNQSSANKTQA--IAKQPINRGQ 1922
 QY 952 PSPYLPERRQL-----HGSSPTRLRQ-----SKRISDSVSDYDCR-----987
 DB 1923 PKILQKSTFPSSQKDIIDRGAATDEKQNFALNTVTCFSNSSLSSLDIQENNK 1982
 QY 988 -----DGVGVVSDYRNG-----RDLQSSTLSVPEQVMSNHC 1020
 DB 1983 ENBPKEBTPPDQSGPSKPOASGYAPKSFHVEDTVCFSNSSLSSLSIDSEDLLOEC 2042
 QY 1021 SPGSGPHRVD---VIGTRSNPSAPPORNVQGRGTRATGHTYTIISMD-RHRVMD 1076
 DB 2043 ISSAMPKKKPSKLGDNKHSKSP-----RNM-GGILGHDLTLDKDIQRFDSHGLSPD 2095
 QY 1077 HYSDDRDRDCEAAD-----RO-----FYHRSRSTE 1101
 DB 2096 SENFDKATQEGANSVSSHLQAAAACLSRQASSDSBLSLSKSGISLSPPLTPDQE 2155
 QY 1102 QRELL-----ERTTRSRSSERPDYN-----1122
 DB 2156 EKFTNKGPRILKPKGEKSTLETLEKSESKGKGGKVKYKSLITCKVRSNSISQMKQ 2215
 QY 1123 -LMRSPSLMTGSAAPPSPALSRSHPTGSV-----QTSPTSSTFGTRGRGQLPOLPP 1174

```

Db 2216 PLQANPSTSRGTMTHIPGRNNSSTSVSKGPKLTPAKSPSEG-----QTATTSP 2271
QY 1175 KGTLSAMDIBERNQMLNKYKQVAGSDPRLQDYHYSKYSRG-----WDPHRCAD----- 1226
Db 2272 RGAKPSVKSELSVPAQT-----SQIGGSB-----KAPSRGSRDSTPSRPAQOPLSR 2319
QY 1227 TVSTKSDSDVSVAVSTSSSRFSSTSYMSVQSERPGRNPKISVFTSKMQRONGVS 1286
Db 2320 PIOSPGRNISPGRNGISPNKLSQLPRTSPSTASTKSSGSKMS-YTS--PGRQW--S 2374
QY 1287 GKLTSTSGDMCSLEKNDGSDSDTAVGALCTSGKRRSSIGAKWVAIVGLSRKSRSA 1346
Db 2375 QMLTQTGLSKWASSIPRSESASKGLNQWNGGANK-----VELSRMSTK 2423
QY 1347 SOLQSTGGGKKL-----RSTVQRSTSTGLAVMRNMWTRQASRST----- 1388
Db 2424 SSGSRSRSPVLVRSQSTFKAPSPFLRRKLESASFELSPSRPASPTESQATPV 2483
QY 1389 -----DGSNVSYS 1397
Db 2484 LSPSLFDMLSLTHSS 2498

```

RESULT 10

US-08-450-582-7

Sequence 7, Application US/08450582

Patent No. 6114124

GENERAL INFORMATION:

APPLICANT: ALBERTSEN, HANS

APPLICANT: ANAND, RAKESH

APPLICANT: CARLSON, MARY

APPLICANT: GRODEN, JOANNA

APPLICANT: HEDGE, PHILIP J.

APPLICANT: JOSLYN, GREGG

APPLICANT: KINZLER, KENNETH

APPLICANT: MARKHAM, ALEXANDER F.

APPLICANT: NAKAMURA, YUSUKE

APPLICANT: THLIVERKIS, ANDREW

TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC

TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS

NUMBER OF SEQUENCES: 102

CORRESPONDENCE ADDRESS:

ADDRESSEE: Banner & Witcoff, Ltd.

STREET: 1001 G Street, NW

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20001-4598

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/450,582

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/452,655

FILING DATE: 25-MAY-1995

APPLICATION NUMBER: US 08/289,548

FILING DATE: 12-AUG-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/741,940

FILING DATE: 08-AUG-1991

ATTORNEY/AGENT INFORMATION:

NAME: Kagan, Sarah A.

REGISTRATION NUMBER: 32,141

REFERENCE/DOCKET NUMBER: 1107.49964

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-508-9100

TELEFAX: 202-508-9299

INFORMATION FOR SEQ ID NO: 7:

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 2843 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; US-08-450-582-7

```

Query Match 3.1%; Score 258.5; DB 3; Length 2843;

Best Local Similarity 17.9%; Pred No. 4.4e-10;

Matches 304; Conservative 238; Mismatches 618; Indels 535; Gaps 69;

```

QY 30 SHLTERRKIIILAVMDRQKREKQSVLKIKKHAQPTWFFPSGIGITELVNNYLPQOQ 89
Db 1012 NEMDNDGELDTPINYSKYSDQ-----LNSGRQSPQMRWARPKHIE--DEIKQSQ 1065
QY 90 KQNEKEPQTKLHQQFEMTKQVKMGESQ---QORQKGDAPTCGICHTKTFADGCH 146
Db 1066 RQSRNQS-----TTPYVYTESTDDKHLKPOPHFQOE-----CVSPTREGANGSET 1112
QY 147 NCYQCTKFCARCGGRVSLRKNVNVNLCRQKQILTKSGAWFTNSGNTLQPDQKV 206
Db 1113 N-----RVGNHGINQVNSQLC-----QEDYEDDKPTWYSEYSEEBQHEZ-- 1155
QY 207 PRGLNNEAPQKKAHLHQPOFOGAPGDLSPVAVKRAHGLTRQDTIKNGSVKHQTA 266
Db 1156 -----EERTVYSIKYNEKRVDPIDYSL-----XVA 1184
QY 267 SDMPDRKESPSVRDQNRREYQSEEREDYSQVPSDGTWPRSPDYADRRQRPQFYE 326
Db 1185 TDIPSSQKQSPFSKSSGQSKTE-----HMSSESTSTPSNAKRONQLHPSSAQ 1237
QY 327 EPCHLYRDSNRKGRHSKE-----YIVDD-----EDVE 355
Db 1238 SRSGQPKAATCKVSSINQETITQVCVETPTICFSECSLSLSSAEIDIGCMTQTEAD 1297
QY 356 SRDEYERQREBEYQARYSDPNLARYPKVQPYEQEOMRIHAIVSRARHERHSDVSLAN 415
Db 1298 SANTIQAIEIKIGITRSEAEDPVSEVPAVSQHPRTKSSRLQGS-SLSSSARHKAVEPSS 1356
QY 416 AELEDSRISLLMRDP-----SRQSVSERRAAMENQRSYSMERTBAQ----- 459
Db 1357 GAKSPSKGAQTPKSPPEHYVOSTPLMFSECTSVS-----SLDSFESRSIASSVQSPCSG 1412
QY 460 -----GQSSYPORTSNHSPPTPRSPPIPLDRPDMERADSLRQHLDPSSAVRKTREK 513
Db 1413 MVSGIISPSDLFDSPOQTWPPSRKTPPP-----PQTA--QTRR- 1451
QY 514 METMLNDSLSDQSGSRVPPPPRPHSKKGGKMRQVSLSSSEHEL-----ASTPEYTC 568
Db 1452 ---VPKNKAPTAKRES-----GPKQAAVNAVQVRQVLPDADTLHFEATESTPQFSC 1502
QY 569 D-----DVEL-----ESRSVSEKDSQ-KKQKT--SEOG 595
Db 1503 SSSLSALSDEPFIQKDVLRIMPVQENDNGNETSEQKESNENQEKAKETIDSEKD 1562
QY 596 VLSDSN-----TSEROKRMYVYGGHLEEDLEWSEPOIKDSGVDTCSG 639
Db 1563 LLDSDDDDIIEIEECIIISAMPTKSRKAKPAQTASKLPPPVARKPSQI-----PVTK 1616
QY 640 TTLNREHSHSKHPVTWQPSKQDRL--IGRILLN-----KRLKDGVPFSDGAWLGL 690
Db 1617 LLPSONRLQPKH-VSFTPDQDMRPVYCVGEGTPIFNSTATSLDITIEPPFELAAGECV 1675
QY 691 KVVGGKQTESGRLCAPITVKKGSADTVGHLRPGDEVLEWGLLQGAFFEVNIIIE 750
Db 1676 R--GG--AQSG-----EFKRDITPTREG--RSTDEA-----QQCKTSVITPILD 1714
QY 751 SKPEQVELVV-----SRPIDG-----IPRIPDSTHAQLESSESSSFESQKDRSISTV 799
Db 1715 DNKAERGDIILBCINSAMPKSKHFKFRVKIMDQVQ-QASASSAPNKNQDKKKKPT 1773

```

QY 800 SPSPGMLRUVPOPLSQSLIKLWFDKVGHQLIVITLGAOKLPSRDRGRPNVYKLYL 859
 Db 1774 SPVKP-----IPONTETRTV-----RKNADSKNLAERVF 1805
 QY 860 PDRSDKMKRTKTKVKTLEPK-----MNQTFYSPVH-----RR 893
 Db 1806 SDNDSKQMLKNSKDFNKLPNEDRVGSAFADSPHYHTPIEGTPYCFSRNDSLSL 1865
 QY 894 EFERMLEITLMOARVRESEFGLBILILETALDDPHWYKLTQTHDVSSLPFR-- 951
 Db 1866 DFDDDDVLSR-EKAEKAKENKSEAKVTSHTLTSNQSANKTQA--IAQPINRGQ 1922
 QY 952 PSYLPFRQL-----HGSPTRRLQR-----SKRISDSFVSDYDCE-- 987
 Db 1923 PKPILQKSTFPOSSKIDPRGAATDEKLFQFAIENTFVCFSHNSLSLSLSDIQENNK 1982
 QY 988 -----DGVGVSDVYRNG-----RDLQSSLTSLVPEQWSSNHC 1020
 Db 1983 ENPIKETEPDQGEPSKPOAGYAPKSFHVEDTFCFGRNLSLSLSIDSDLLQBC 2042
 QY 1021 SPGSHRVD---VIGETREWSAPPORNVQGHRTGTRATGHYNTISMD-RHRVMD 1076
 Db 2043 ISSAMPKKKPSRLKGNKHSK-----RNM-GGILGEDLTLDLKDIOQPDSEGLSPD 2095
 QY 1077 HYSDDRDCEAD-----RQ-----PHRSRSTZ 1101
 Db 2096 SENFDWKAIOEGANSIVSSLHQAACLSQASDSDSLSLKSLGSLGSLGSLGSLGSL 2155
 QY 1102 QRPULL-----ERTTRESRSPDTN----- 1122
 Db 2156 EKFTSNKGPRIKPKESKSTLTKLIESKGIKGGKVKYKLTGKVRNBSISQMKQ 2215
 QY 1123 -LMSRPSLMTGRSAPPALSRHPRTGSV-----QTSPTSTGTGRGRQLPLPP 1174
 Db 2216 PLQANPISGRTHIIPGRNSSSTSPVSKKGPPLKTPASKSPSEG-----QTATTSP 2271
 QY 1175 KGTLSRMDIERNRQMLNKYQVAGSDPRLQDYHYSKVRG---WDFRGAD----- 1226
 Db 2272 RGAKPSVKSLSVART-----SQIGSS-----KAPSRSGSRDSTPSRPAQPLSR 2319
 QY 1227 TVSTKSDSDVSDVSAVSTSSARFSTSYMSVQSRPGRNKLIVFTSMQRONGVS 1286
 Db 2320 PIQSPGRNISPGRNGISPPNKLSQLPRTSSPTASTKSGSGKMS-YTS--PGRQ--S 2374
 QY 1287 GKMLTSTISGDMCLKNDGSDTAVGALGTSGKKRESSIGAKNVAIVGLSRKRSRA 1346
 Db 2375 QQNLTKGTGLSKNASSIPRESASKGLNOMNGANK-----VELSEWSTK 2423
 QY 1347 SOLSQTEGGGKL-----RSTVQSTETGLAVENRNMTRQASRST----- 1388
 Db 2424 SSGSEDRSRPVLVROSTFIKEAPSTLRRKLEESASFESLSPSRPASPTRSQATPV 2483
 QY 1389 -----DGSNNYSYSS 1397
 Db 2484 LSPSLPDMSLSTHSS 2498

RESULT 11
 US-08-821-355A-7
 ; Sequence 7, Application US/08821355A
 ; Patent No. 5851775
 ; GENERAL INFORMATION:
 ; APPLICANT: Barker, Nick
 ; APPLICANT: Clevera, Hans
 ; APPLICANT: Korinek, Vladimir
 ; APPLICANT: Morin, Patrice
 ; APPLICANT: Kinzler, Kenneth
 ; APPLICANT: Vogelstein, Bert
 ; APPLICANT: Sparks, Andrew
 ; TITLE OF INVENTION: Beta Catenin, TCF-4, and APC
 ; TITLE OF INVENTION: Interact to Prevent Cancer
 ; NUMBER OF SEQUENCES: 11

; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Banner & Witcoff, Ltd.
 ; STREET: 1001 G Street, N.W.
 ; CITY: Washington
 ; STATE: DC
 ; COUNTRY: USA
 ; ZIP: 20001
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSeq for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/821,355A
 ; FILING DATE: 20-MAR-1997
 ; CLASSIFICATION: 514
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Kagan, Sarah A
 ; REGISTRATION NUMBER: 32,145
 ; REFERENCE/DOCKET NUMBER: 1107.05064
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 202-508-9100
 ; TELEFAX: 202-508-9299
 ; TELEX: 97430 RMB UT
 ; INFORMATION FOR SEQ ID NO: 7:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2973 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: No. 5851775e
 ; US-08-821-355A-7

Query Match 3.1%; Score 258.5; DB 2; Length 2973;
 Best Local Similarity 17.9%; Pred. No. 4.7e-10;
 Matches 304; Conservative 238; Mismatches 618; Indels 535; Gaps 69;
 QY 30 SLLTEERKILLAVMDROKKEKEKQSVLKKEHKAQTPQWPFPSGITYLNNVLOPQ 89
 Db 1012 NHMDNDGELDTPTINSLKTSDEQ-----LNSGRQSPSONERWAPKHIIE--DEIKQSE 1065
 QY 90 KQNEKEPQTKLHQFEMTYKEQVKNGESQ---QOEKQGDAPTGGICHTKTKFADGCGH 146
 Db 1066 RQSENGS-----TTPVVTSTDDKHLKFPHPGQGE-----CVSPYRSRANGSET 1112
 QY 147 NCSYQTKFCARCGRVSLRSNKVMVCMCRKQOQILITSGAWFNSSGNTLQQPDQKV 206
 Db 1113 N-----RVGSHNGINQNVQSGLC-----QEDDYEDDKFTNYSRYSEEQHEE-- 1155
 QY 207 PRGLNEEAPQEKKAKLHEQPOQAGDLSVPAVEKRAHGLTRQUTINGSGVGHQIA 266
 Db 1156 -----EERTNYSIKYNEEKHVDQPIDYSL-----KYA 1184
 QY 267 SDMPDRKSPSVSRDQNRRYEQSEREDYSQVPSDGTMPSPSPDYADRSQREPOFYE 326
 Db 1185 TDIPSSQKQSPFSKSSSGSSKTE-----HMSSESTSTPSSNAKQNLHPSSAQ 1237
 QY 327 EPGLNYSNRNRGRHRSKE-----YIVDD-----EDVE 355
 Db 1238 SRSGQPKAATCKVSSINQETIQTICVEDTPICFSRCSLSLSLSAEDTICGQNTQEAD 1297
 QY 356 SRDEYERQREHEVQARYSDPNLARYPKQPYEQMIAHVSARHRRHSYSLAN 415
 Db 1298 SANTLQIAETKEKIGTRSAEDPVSEVPAVSQHPRTKSRSLQGB-SLSSSASAKHKAFFSS 1356
 QY 416 AKLEDKRSILLRMDRP-----SKQSVSRRAAMENORYSYMERTEAQ----- 459
 Db 1357 GAKSPKSGAQTPKSPPEHVQETPLMFSECTSVS-----SLDSFERSSTASSVQSPCSG 1412
 QY 460 -----GQSSYPORTSNHSPPTPRRSPILDRPDMRADSLRKQHLDPSSAVRKTREK 513


```

QY 147 NCSYCTKFCARCGGRVLSRKNMVMVCMCRKQOELTKSGMFTNSGNTLQDPQKV 206
Db 1113 N-----RVGSHGINQNVQSLC-----QEDDYEDDKPTNYERYSEBEEH-- 1155
QY 207 PRGLRNEEAPQEKAKLHEBPQPGAPGLDVPVAVEKGRAGHGLTRQDTTKNGSVYHJOIA 266
Db 1156 -----EERPTNYIKINBEKHVDQPIDYSL-----KYA 1184
QY 267 SOWPSDRKSPSVSRDQNRERYEQSEREDYSQVPSDGTWPSRPSYADRRSOREPQFYE 326
Db 1185 TDIPSSQKQSPFSKSSQSKTE-----HMSSESNTSTPSSNAKQNLHPSSAQ 1237
QY 327 EPGHLNYSRNRGHHSHKE-----YIVDD-----EDVE 355
Db 1238 SRSGOPKAATCKVSSINOETIQYCVEDTPICFSRCSLSLSAEDBIGCNOQTQEDAD 1297
QY 356 SRDEYERORREERYQARYSDPNLARYPVKQPYEQBOMRIHARVBARHERHSDVSLAN 415
Db 1298 SANTQIAIEIKIGTRSAEDPVSVFVAVSQHPRITKSSRLQGS-SLSSSARHQAVEFSS 1356
QY 416 AELEDSRISLLMRDP-----SRQSVSERRAAMENQRSYSMERTREAO----- 459
Db 1357 GAKSPKSGAQTPKSPPEHYVQETPLMFSRCTSVS-----SLDSFESRSIASSVQSEPCSG 1412
QY 460 -----GQSSYPORTNSHSPPTPRRSPILDRPDMRRADSLRKQHLDPSSAVKTKRREK 513
Db 1413 MVSGLTSPDLPSQGTWPPSPRSKTPPP-----PQTA--QTKRE- 1451
QY 514 METMLANDSLSDQSVSRPPPPRPHKSKKGGRQVSLSSSEEL-----ASTPEYTC 568
Db 1452 ---VPKNKAPTAEKRES-----GPKQAAVNAAVQVQVFLDADTLIHFATESTPDGFSC 1502
QY 569 D-----DVEL-----ESESVSERKGDQ-KGKRT--SEQ 595.
Db 1503 SSSLSALSDEPTIQDVELRIMPVQENDNGNETSEQPKESNENQKEAKTIDSEKO 1562
QY 596 VLSDSN-----TRSRQKRMYYGHSLEEDLEWSEPOIKQSGVDTCS 639
Db 1563 LLDSDSDDDITELRECIISAMPTKSKKAKKPAQTASKLPPPVARKPSQL-----PVYK 1616
QY 640 TTLAREHSHDKHPTVWQSKDQDL--IGRILLN-----KRLKQGVPRDSGMLGL 690
Db 1617 LLPSONRLOPQKH-VSFTPGDDMPRYCVEGTPIFSTATSLDITSPFNELAAAGEV 1675
QY 691 KVVGGKMTSGRLCAFTTKVKGSLADTVGHLRPGDEVLEWNGRLQAGATFEVYNNIL 750
Db 1676 R--GG--AQSG-----EPEKDTIPEG--RSTDEA-----QGGKTSVTIPELD 1714
QY 751 SKPEQVELV-----SRPIGD-----IPRIPDTHAQLLESSSSSFBSQKMDRPSISVT 799
Db 1715 DNKAERGDILAEICINSAMPKGSKHPKPVVKIMDQVQ-QASASSAPAKNQLDKKKKPT 1773
QY 800 SPMSPCMLRDPVQFLSQSLIKLWFDKVGHLVITLGAKDLPSREDGRPNPVKIYFL 859
Db 1774 SPVKP-----IPONTYRTRV-----RKNADSKNLNAERV 1805
QY 860 PDRSDKNKRTKVKKTLKPK-----MNQTFIYSPVH-----RR 893
Db 1806 SDNKSQKQNLQNSQDNDKLPNEDVRGSPAPDSPHYTFIEGTPYCFSRDLSL 1865
QY 894 EFERMLRITLMDQARVRESEFLGEILLTELTALLDDEPHYKLTQTHDVSSLPFR-- 951
Db 1866 DFDDDDVDLSR-EKALRKAKENKESAKVTSHTLTSNQOSANKTQA--IAQPINRQ 1922
QY 952 PSYLPFRQL-----HGSPTRRLOR-----SKRISDSVSDYDCE----- 987
Db 1923 PKPILQKSTFPOSSKDIPDRGAATDEKLNQFAIENTPVCFSHNSLSLSLSDIQENNK 1982
QY 988 -----DGVGVSVDYRHNG-----RDLQSTLSVPEQVMSSNHC 1020
Db 1983 ENBPKEPTEPPSQGBSPKQASGYAPKSFHVEDTPVCFSRNSLSLSLSDSDLLQEC 2042

```

```

QY 1021 SPSCGSHRYD---VLGRTRMSPPSAPPQPNVBOGHRGTRACHNTNTISMD-RHRVMD 1076
Db 2043 ISSAMPKTKKPKSRKAGDNKHEKSP-----RNM-GGILGEDLTDLKDIOQRPDSHGLSPD 2095
QY 1077 HYSSDRDRDCBAAD-----RO-----PYHRSRSTE 1101
Db 2096 SENFWKALQOGANSIVLSLHQAAAAACLSRQASDSLSILKSGISLGSPFLTPDOE 2155
QY 1102 ORPLL-----ERTTTSRSRSPDTN----- 1122
Db 2156 EKPFSTNKGPRILKPKGKSTLTETKLESKGIKQKVKYKSLITKGVNSSEISQMKQ 2215
QY 1123 -LARSWPSLWCRSAPPSPALSRSPRTGSV-----QTSPPSTPGTGRGROLPOLPP 1174
Db 2216 PLOANPISIRGCTMIHIPGVNNSSTSPVSKGPPPLKTPASKSPFSG-----QTATTSF 2271
QY 1175 KZTLERSAMDIEERNKQMLNKYKOVAGSDPRLQDTYSKYRSG--WDPHRGAD----- 1226
Db 2272 RQAKPVSSELSFVAKOT-----SQIGSS-----KAPSGSRDSTPSRPAQOPLSR 2319
QY 1227 TVSTKSDSDVSDVSAVGRKTSASRFSSTSTMSVQSERPRGRNKISVFTSKNQRMGYS 1286
Db 2320 PIQSPGRNSISPGRNGI8PPNKLSQLPRTSSPSTASTKSSGSGKWS-YTS--PGRQM--S 2374
QY 1287 GKRLTSTISIDMCLEKNDCSQSDTAVGALGTSGKKRSSIGAKMVAIVGLSRKSEA 1346
Db 2375 QONLQTKQTLGSLNASSIPRSESASRGLNQMNGGANKK-----VELSRMSSTK 2423
QY 1347 SQLSOTEGGKKL-----RSTVORSTETGLAVEMENMTQASRST----- 1388
Db 2424 SSGSBSRSPVLRQSTFIKAPSTLURKLESASFESLSPSRPASPTRSQATPV 2483
QY 1389 -----DGSMSYS 1397
Db 2484 LSPSLPDMSLSTHSS 2498

RESULT 13
US-09-136-605-7
; Sequence 7, Application US/09136605A
; Patent No. 6140052
; GENERAL INFORMATION:
; APPLICANT: He, Tong-Chuan
; APPLICANT: Kinzler, Kenneth
; APPLICANT: Vogelstein, Bert
; TITLE OF INVENTION: Beta Catenin, TCF-4, and APC Interact to
; FILE REFERENCE: 1107.75741
; CURRENT APPLICATION NUMBER: US/09/136,605A
; EARLIER FILING DATE: 1998-08-20
; EARLIER APPLICATION NUMBER: 08/821,355
; EARLIER FILING DATE: 1997-03-20
; EARLIER APPLICATION NUMBER: 09/003,687
; EARLIER FILING DATE: 1998-01-06
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 2973
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-136-605-7

Query Match 3.1%; Score 258.5; DB 3; Length 2973;
Best Local Similarity 17.9%; Pred. No. 4.7e-10;
Matches 304; Conservative 238; Mismatches 618; Indels 535; Gaps 69;

QY 30 SHLTREERKILAVMDRQKEEKESQVLKIKERHKAQPTWPPFSGITELVNNVLPQ 89
Db 1012 NEMDDNDGELDTPIYSLKYSDEQ-----LNSGRQSPSOMERWARPXHIIE--DEIKQSQ 1065
QY 90 KQBNKEPQTKLHQEPQEMTKQVKQMGESQO---QOQKGDAPTGGICHTKTPADGGH 146
Db 1066 ROSRQNS-----TTYPVYTESTDDKHLKFPQFQOE-----CVSPYRSGANGSET 1112

```


REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107.035574
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2843 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-741-940-2

Query Match
Best Local Similarity 18.0%; Score 250.5; DB 1; Length 2843;
Matches 305; Conservative 235; Mismatches 620; Indels 535; Gaps 70;

30 SHLTREERKILLAVMDRQKEKEKESVLIKKEHKAQPTQWPFPGITELVNVYLOPOQ 89
1012 NHMDNDGELDTFTINSLKYSDEQ-----LNSGRQSPQNERWARPXHLIE--DEIKQSQ 1065
90 KQPKNEKEPQKLHQPEMTEKQVKKMGESQ---QOQKQGDAPTGGICHTKTFADGCGH 146
1066 QKSENGS-----TTPVVTESDDKHLKFPHPGQOE-----CVSPYRSGANSET 1112
147 NCYCCTKFCARCGGVSLRSNVMVCMCRKQBIILTKSGAWFNNGENTLQOPDKV 206
1113 N-----RVGSNHNQNVQSLC-----QEDDYEDDKFTNYSERYSEBQHEE-- 1155
207 PRGLRNEEAPQEKAKLHQPQPGAGDLSVPAVEKRAHGLTRODTIKNGSVKHOA 266
1156 -----EERPTNYSIKYNEKRVHDQPIDYSL-----KYA 1184
267 SDMPDRKSPSVSRQNRQRYEQREREDYSQVPSDGTMPRSPDYADRRSQREPOFYE 326
1185 TDIPSSQKQSPSKSSQSSQSKTE-----HMSSESTSTPSSNAKRONQHLFSSAQ 1237
327 EPHLYRDSNRGRHRSKE-----YIVDDDV-----ESRDEYERQRREEYQ 370
1238 SRSGQPKAATCKVSSINQETIQYCVEDTPICFSRCSLSLSSAEDDEIGCMTQTEAD 1297
371 A-----RYSDPDLNARYPVKQPYEQWRIHARVSRARHRRHSYSLAN 415
1298 SANTLQIAIKGIGTRSAEDPVSEVPAVSQHPRTKSSRLQGS-SLSSSEARHKAVERFS 1356
416 AELEDKRIILLRMDP-----SRQSVSEKRAAMQNSYSMERTREAO----- 459
1357 GAKSPKSGAQTPKSPPEHYVQETPLMFSECTSVS-----SLDSFERSSTASVQSEPCSG 1412
460 -----GQSSYPQRTSNHGPPTPRRSPILDRPDMRRADSLRQKHLDPSAVKRTYREK 513
1413 MVSGIISPLDPSQGTMPPSRSKTPPP-----PQTA--QTKRE- 1451
514 METMLANDLSLSSQSESVPRPHKSKGQKQROVLSLSSSEEL-----ASTPYTSC 568
1452 ---VPKNKAPTAEKRES-----GPKQAAVNAAVQVQLPDADTLHLFATESTPPGFSC 1502
569 D-----DVEL-----ESESVSSEKGDQO-KGKRKT--SBQG 595
1503 SSSLALSLEDPFIQKDVRLRIMPVQENDNGNETSEBQPKSNENQEKAEKTIIDSEKD 1562
596 VLSDSN-----TRSERQKRMITYGGHSLSEEDLEWSEFPQKDSGVPTCSS 639
1563 LLDSDDDDIIELEECIIISAMPTKSRKGGKPAQTASKLPPPVARKPSQL-----PVYK 1616
640 TTLNEESHSDKHUPWQPSKDGRL--IGRIILN-----XRLKDGSVPRDSGMGL 690
1617 LLPSONLQPKH-VSPTEGDMFVVCVETPTNFSTATSLDSTIESPPNELAAGBV 1675
691 KVVGGKMTESGRCLCAFTTKVKKGLADTVCHLRPGDSEVLEWNGRLIQATFEFVYNIIE 750
1676 R--GG--AQSG-----EPEKDDTIPTEG--RSTDEA-----QGGTKSSVTIPELD 1714

751 SKPEPOVELV-----SRPIGD-----IPRIPDSTHQALESSESSSFFBQKMDRPSISVT 799
1715 DNKAEGDILAEICINSAMPKGRSHKPRFKKIMDQVQ-QASASSAPKNGOLDKCKKCT 1773
800 SPMSQGLRDVQFLSGQLSILKMPDKVGHQILVILCAKDLPSREDGKPRPVKVIYFL 859
1774 SPVKP-----IPQNTYTRV-----RNVADSKNNLAEVFP 1805
860 PDRSDKRRRTKTVKTLLEPK-----WNOTTYISFVH-----RR 893
1806 SDWKDSKQNLKNSKDFNDKLPNNEDRVGRSFAFDSPHHYTPIGTFYCFPSRNDLSL 1865
894 EFRERMLEITLMDQARVRESEFFGEILIELEALLDDEPHYKLTQDIDVSLPLPR-- 951
1866 DFDDDDVDLSR-EKAELEKAKENKESAKVTSHTELTSNQOQANKTQA--IAQPTINRQ 1922
952 PGPYLPERRQL-----HGESPTRLQR-----SKRISDSRVSVDCE-- 987
1923 PFILOKQSTFPQSSKNDIPDRGAATDEKLONTAIENTVCFPSHNSLSLSLSDIDQNNK 1982
988 -----DGVGVVSDYRNG-----EDLOSSTLSVPEQVMSNHC 1020
1983 ENERIKETEPDPSQGEPSKQASGYAPKSPHVEDTPVCFPSRNSLSLSLSDIDQJQC 2042
1021 SPGSGFHRVD---VIGTRSHSPSPAPPQPNVTEQGRGTRATGHYNTISMD-RRKMD 1076
2043 ISSAMPKPKKPSRLAGDNEKHSF-----RNN-OGILGEDLTLDLNDIQRPSEHGLSPD 2095
1077 HYSSDRDRDCEAAD-----RO-----PYHRSRST 1101
2096 SENFPAKIQEANSIVSLHQAACLSRQASDSDSILSKGILSGSPHUTPDOR 2155
1102 QRPLL-----ERTTTSRSESRPDYN----- 1122
2156 EKPFSTNKGPRILKPKCKSTLTETKIESKGIKGGKVKYKSLITCKVESNEISQMKQ 2215
1123 -LWRSHPLMTGRAPPSPALSRSHPTQSV-----QTSPPSTGTQRRGRQLPOLPP 1174
2216 PLQANPISIRGRTMIHIFGVNNSSTSFVSKGPPPLATPASKSPSEG---QTATTSP 2271
1175 KTWLERSAMDIERNRQMLNKYKQVAGSDPRLQDYHSEKTSG---WDPHRGAD-- 1226
2272 RGAKPSVSKSELSFVARQT-----SOIGSS-----KAPSRSGSRDSTPSRPAQOPLSR 2319
1227 TVSTKSDSDVDVAVSRTSSASRFSSTYSMSVQSERPGRKISVFTSKQNRQVGS 1286
2320 PIQSPGNSISPGRNGISPPNKLSQLPTSSPSTASTKSSGSGKMS-YTS--PGRCM--S 2374
1287 GKNLTKSTISIGDMCSLEKNDGOSDTAVGALGTQKTKRSGSIGAKWVAIVGLSRKSA 1346
2375 QOQLATQTLGLSKNASSIPRSESASKLQNMNNGANGANK-----VELSRMSSTK 2423
1347 SQLSQTTEGGKKL-----RSTVORSTETGLAVEMENMTQASREST----- 1388
2424 SSGSESDRERFVLVRQSTFIKEAPSPULRKLRESASFESLPSRSPASPTKSOAQTV 2483
1389 -----DGSMSYS 1397
2484 LSPSLPDMSLSTHSS 2498

RESULT 15
US-08-289-548A-2
Sequence 2, Application US/08289548A
Patent No. 5648212
GENERAL INFORMATION:
APPLICANT: ALBERTSEN, HANS
APPLICANT: ANAND, RAKESH
APPLICANT: CARLSON, MARY
APPLICANT: GRODEN, JOANNA
APPLICANT: HEDGE, PHILIP J.
APPLICANT: JOSLYN, GEOFF
APPLICANT: KINZLER, KENNETH

APPLICANT: MARKHAM, ALEXANDER F.
 APPLICANT: NAKAMURA, YUSUKE
 APPLICANT: THLIVERIS, ANDREW
 TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
 TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
 NUMBER OF SEQUENCES: 102
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Banner & Allegretti, LTD
 STREET: 1001 G Street, NW
 CITY: Washington
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20001-4598
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/289,548A
 FILING DATE: 12-AUG-1994
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Kagan, Sarah A.
 REGISTRATION NUMBER: 32,141
 REFERENCE/DOCKET NUMBER: 1107.46943
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202-508-9100
 TELEFAX: 202-508-9299
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2843 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-289-548A-2

Query Match 3.0%; Score 250.5; DB 1; Length 2843;
 Best Local Similarity 18.0%; Pred. No. 1.8e-09;
 Matches 305; Conservative 235; Mismatches 620; Indels 535; Gaps 70;

30 SHUTEERKILAMDRQKKEKESVLSKVKHKAQTPWFFSGITELVNVVLPQQ 89
 1012 NHDDNDGELDTPIYSLKYSDEQ---LNSGRQSPQNERWARPKH112--DEIKQSE 1065
 90 KQNEKEPQTKLHQQFEMYEQVKQMGERSQ---QEQKGDAPTGCICHTKTFADGCGH 146
 1066 QQRNQS-----TTTPVYTESTDDKHLKQPHFGQE-----CVSPYRSGANGSET 1112
 147 NCSYCQTKFCARCGRVSLASNKVMVYCNLCRQKQILTKSGAMFYNSGNTLQOPDQKV 206
 1113 N-----RVGSNHGINQNVQSLC-----QEDDYEDDKPTNYSEYSEHEQHEE-- 1155
 207 PGRLENEAPQEKAKLHQEQPQAGDLSPVAVEKGRAHGLTRODTIKNGSGVHQTA 266
 1156 -----EERTNYSIKYNEKRVHDQIDYSL-----KYA 1184
 267 SDMPDRKSPSVSRDONRYEQESEEREDYSQVPSDGTWPRSPDYADRRSQREPOFYE 326
 1185 TDIPSSQKQFSPKSSGSSQSKTE-----HMSSESTNTSPSSNAKQNLHSSAQ 1237
 327 EPGLNRYDSNRGRHRSKE-----YIVDDSDV-----ESRDEYERORREBEYQ 370
 1238 SRSQPKAATKCVSSINQETIQYCVEDTPICFSRCSLSLSLSSAEDEIGCQNTQEAD 1297
 371 A-----RYRSDPNLARYPVKQPYEQPQYEEQMRJHAEVSRARHRRHSIDSVLAN 415
 1298 SANTLQIAEIKGKIGTSAEDPVSEVPASVQHPRTKSSRLQGS--SLSESARHKAVEFPS 1356
 416 AELEDSRISLLEMDRP-----SRQSRSEERRAAMENQORSYSMEETREAQ----- 459
 1357 GAKSPSKSGAQTPKSPPEHYVQETFLMFSRCTSVS-----SLDSFESRSIASSVQSEPCSG 1412

460 -----GQSSYPORTSNHSPPTPRRSPPIPLDRPDMRRADSLAKQHLDPSSAVRNTKREK 513
 1413 MVSGIISPDLPDPSGQTPMPPSRKTPPP-----PPQTA--QTKEE- 1451
 514 METMLRNDLSLSDQSESVRPPPHKSKKGGKQVSLSSSEHEL-----ASTPEYTS 568
 1452 ---VPNKAPTAKRES-----GPKQAAVNAVQVQVLPDADTLLHFAFESTDPGSC 1502
 569 D-----DYEL-----ESESVSSEKGDQ--KGRKT--SRQG 595
 1503 SSSLALSILDEPFIQKQVLELIMPVQENDNGNETESQPKESNENQEKAKETIDSEK 1562
 596 VLSDN-----TREROKKMYGCHSLEDELEWSHQPQIKDQGVDTGSS 639
 1563 LLDDDDDDIEILERCIISAMPTKSRKKGKPAQATASKLPPVAPKPSGL-----PVYK 1616
 640 TTLNEESHSDKHVPTWQPSKGDDEL--IGRILL-----KRLKDSVPRDSCAMGL 690
 1617 LPSQRRLQPKH-VSFTPGDDMPRYCVGTPPIINFSTATSLSDUTISPPNELAAGEV 1675
 691 KVVGGKMTESGRLCAFITKVKKGLADTVGHLRPGDQVLENGRLLQAGATFEVYNIILE 750
 1676 R--GG--AQSG-----EFKRDITPIEG--RSTDEA-----QGGKTSSTVITPELD 1714
 751 SKPEQVELV-----SRPGB-----IPRPDTHAQLESSESSSFOSQWDRPSISVT 799
 1715 DNKAEEGDLAEICINSAMPKGSKSHKPPRVKIMDQVQ--QASASSAPKKNQLDGKKKPT 1773
 800 SPMSRGLRDVQPLSGQLSIKLMFQKVLQVILGAKDLPREDGPRNPYKVIYPL 859
 1774 SPVK-----IPONTYETRV-----RKNADSKNMLAERV 1805
 860 PDRSDKNRRTKTVKTKLEPK-----WNQTFIYSPVH-----RR 893
 1806 SDNDSKKNLKNKSKDFNDKLPNNEDRVGSPAFDPSPHYTPIRGTPYCFKSNDSLSL 1865
 894 EFRERMLEITLWDQARVRESESEFGEILIELEALLDDDEPHYKLTQHDVSSLPFR-- 951
 1866 DFDDDDVDLSR-EKAEIRKAKENKESEAKVTSHTELTSTNQSQANKTQA--IAQPINRQ 1922
 952 PSPYLPRRL-----HGESPTRLQR-----SKRLSDSVSDYDCE--- 987
 1923 PXPILQKQTFPQSSKDIPDGAATDEKQNTFAIENTVCTFHNSSLSSLSLIDQNNK 1982
 988 -----DGVGVVSDYRNG-----RDLQSTLTSVPPQVMSNH 1020
 1983 ENEPKETEPPDSQGEPSKPQASGVAPKSFHVEDTPVCFSRNSSLSSLSIDEDDILQEC 2042
 1021 SPSSGSHRVD---VIGRTSMSPSAPPPOPNVEQGHRTGTRATGHYNTTISRM-D-RHRVMD 1076
 2043 ISSAMPKPKKPKSLKGDNEKHSF-----RNM--GGILGEDLTLDLKDIOQRPDSHGLSPD 2095
 1077 HYSSDRDRDCEAAD-----RQ-----PYHRSRST 1101
 2096 SENFDKAIQORGANSIVSSLHQAACLSRQASDSDLSILKSGISLGSFHLTPDQ 2155
 1102 QRPL-----ERTTTRSRSERDTN----- 1122
 2156 EKPFTSNKPRILKPKGEKSTLETKKIESKGIKGGKVKVKSILTKGVRSNEISQMQ 2215
 1123 -LMSNPMLTGRSAPPSPALSRSHPTQSV-----QTSFSPSTPGTRGRQLPOLPP 1174
 2216 PLOANPISIRGRTIHIHFGVNRNSSLSTSPVSKGPKPLATPASKSPSEB---QTATTSF 2271
 1175 KGTLESAMDIEERNQMLNKYKQVAGSDPRLQDYHSHKYSB---WDPRGAD----- 1226
 2272 RGAKPSVKSELSFVAKQT-----SQIGGS-----KAPSRSGSRDSTFSPAPQPLSR 2319
 1227 TVSTKSSDDVSDVSAVSRSTSGASRPSSTYSVQSERPRGNKIKISVFTSKMQRQMGVS 1286
 2320 PQOSPORNISFORNGISIPPNKLSQLPRTSSPSSTASTKSSGSKMS-YTS--PGRM--S 2374
 1287 GKNLTSTISGDMCSLEKNKDGSDTAVGALGTSGKRRSRISGAKQWVAIVGLSRKRSRA 1346

Db	2375	QONLTKOTGLSKNASSIPRESASKGLNQNNGANKK-----VELSRMSXK	2423
QY	1347	SQLSQTEGGKKL-----RSTVORSTETGLAVEMRNWMTROASREST-----	1388
Db	2424	SSGSRDRSRPVLVRQSTFIKEAPFTLRKLESASFESLSPSRPPASPTRSQATPV	2483
QY	1389	-----DGSNMNYSYSS	1397
Db	2484	LSPSLPDMSLSTHSS	2498

Search completed: December 12, 2003, 07:21:57
 Job time : 63 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 11, 2003, 23:10:57 ; Search time 76 Seconds
(without alignments)
3320.727 Million cell updates/sec

Title: US-09-617-099b-1

Perfect score: 8285

Sequence: 1 NSAPLGRGRPAPTAAAGP.....TTRAGQSLESSTGTPSYRS 1590

Scoring table: BLOSUM62

Gapop 10.0 , Gapert 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A Geneseq_19Jun03.*
1: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8285	100.0	1590	22 AAB73488	Mouse Rim2, a nove
2	2215	26.7	427	23 AAB63531	KIAA 0751 PDZ doma
3	1586.5	19.1	374	19 AAW29640	Human secreted pro
4	1015.5	12.3	237	21 AAB34848	Human secreted pro
5	1010.5	12.2	237	21 AAB34847	Gene 46 human secr
6	989	11.9	2464	22 AAB63174	Drosophila melanog
7	624	7.5	134	21 AAB34849	Gene 46 human secr
8	569.5	6.9	158	22 ABB15089	Human nervous syst
9	566	6.8	128	21 AAB34850	Human secreted pro

10	454	5.6	92	23 AAB63532	KIAA 0751 PDZ doma
11	420	5.1	126	21 AAB40725	Human ORFX ORF489
12	417	5.0	82	23 AAB63179	KIAA 0751 PDZ doma
13	417	5.0	86	22 AAB55848	PDZ encoded domain
14	417	5.0	86	22 AAB57639	PDZ domain. Homo
15	417	5.0	86	22 AAB58049	PDZ domain. Homo
16	417	5.0	86	23 AAB705314	PDZ domain protein
17	307	3.7	1898	20 AAY30795	A human trichohyal
18	295	3.6	3111	22 ABB60327	Drosophila melanog
19	291.5	3.5	2951	22 ABB60291	Drosophila melanog
20	290.5	3.5	5533	22 ABB65772	Drosophila melanog
21	290.5	3.5	5560	22 ABB71160	Drosophila melanog
22	281	3.4	1331	22 ABB66948	Drosophila melanog
23	279	3.4	3664	24 ABR47592	Breast cancer asso
24	278.5	3.4	910	24 ABR47604	Drosophila melanog
25	278.5	3.4	1333	22 ABB66754	Drosophila melanog
26	278.5	3.4	1430	22 ABB58724	Drosophila melanog
27	276	3.3	1833	22 ABB71141	Drosophila melanog
28	274.5	3.3	895	23 ABB69569	Human polypeptide
29	270.5	3.3	2865	22 ABB48336	Human liver peptid
30	270.5	3.3	2665	22 ABB28314	Human peptide #965
31	270.5	3.3	2665	22 ABB33490	Peptide #996 encod
32	270.5	3.3	2665	22 ABB18950	Protein #949 encod
33	270.5	3.3	2665	22 AAM54270	Human brain expres
34	270.5	3.3	2665	22 AAM66665	Human bone marrow
35	270.5	3.3	2665	22 AAM14533	Peptide #967 encod
36	270.5	3.3	2665	22 AAM26950	Peptide #987 encod
37	270.5	3.3	2665	22 AAM02259	Peptide #941 encod
38	270.5	3.3	2665	23 ABB36319	Human peptide enco
39	270.5	3.3	3266	21 AAB42491	Human ORFX ORF2255
40	268.5	3.2	3021	24 ABB75873	Human secretory po
41	267	3.2	2296	22 ABB59050	Drosophila melanog
42	266.5	3.2	1554	23 ABB97332	Novel human protei
43	265	3.2	117	22 ABB63166	Drosophila melanog
44	261.5	3.2	1479	22 ABB70936	Drosophila melanog
45	261.5	3.2	1501	23 ABB69798	Human polypeptide

ALIGNMENTS

RESULT 1
AAB73488
ID AAB73488 standard; Protein; 1590 AA.
XX
AC AAB73488;
XX
DT 25-JUL-2001 (first entry)
XX
DE Mouse Rim2, a novel isoform of Rim.

XX
KW Mouse, murine; Rim2; Rim isoform; Rab3-interacting molecule;
KW Rab3-dependent intracellular vesicle fusion regulator;
KW GTP/GTP exchange factor; CAMK-GRFII; neuron; endocrine tissue;
KW exocytosis; hormone secretion; neurotransmitter release;
KW presynaptic membrane; neuropathy; central nervous system disorder;
KW brain disorder; secretion disorder; pituitary gland; hypothalamus;
KW pancreatic beta-cell; diabetes; parotid gland; zinc finger; PDZ domain;
KW C2 domain.
XX
OS Mus musculus.
XX
XX EPI090986-A1.
XX
PD 11-APR-2001.
XX
PF 31-JUL-2000; 2000EP-0116148.
XX
PR 08-OCT-1999; 99JP-0288372.
XX
PA (JCEP-) JCR PHARM CO LTD.
XX
FA (SEIN/) SEINO S.

PA (GEM) GENETICS INST INC.
 XX Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Werberg D;
 PI Racie LA, Spaulding V, Treacy M;
 XX WPI; 1998-414101/35.
 XX N-PSDB; AAV40485.
 XX New isolated polynucleotides and secreted proteins - obtained from
 PT human foetal kidney, adult testes, placenta, adult brain and foetal
 PT brain cDNA libraries
 XX Claim 22; Page 71-72; 104pp; English.
 XX This is the amino acid sequence of novel human secreted protein
 CC C0618_1, as predicted from human adult brain cDNA clone C0618_1
 CC (see AAV40485). C0618_1 cDNA was isolated using methods which are
 CC selective for cDNAs encoding secreted proteins, and was identified
 CC as encoding a secreted or transmembrane protein on the basis of
 CC computer analysis of the amino acid sequence of the encoded
 CC protein. C0618_1 shows some sequence similarity to rat Rim, a
 CC putative Rab3 effector. Novel secreted proteins (see AAV29636-45)
 CC have been identified in human foetal kidney, adult testis, placenta,
 CC adult brain and foetal brain, and can be obtained by expression in
 CC recombinant host cells. The polypeptides may have e.g. nutritional
 CC activity, cytokine and cell proliferation or differentiation
 CC activity, immunostimulant or immunosuppressive activity,
 CC haematopoiesis regulating activity, tissue growth activity,
 CC activin/inhibin activity, chemotactic or chemokinetic activity,
 CC haemostatic or thrombolytic activity, receptor/ligand activity,
 CC antiinflammatory activity, cadherin or tumour invasion suppressor
 CC activity, tumour inhibition activity, or other activities.
 XX Sequence 374 AA;
 SQ
 Query Match 19.1%; Score 1586.5; DB 19; Length 374;
 Best Local Similarity 85.8%; Pred. No. 7.1e-103;
 Matches 321; Conservative 6; Mismatches 6; Indels 41; Gaps 1;
 QY 1258 MSVQSEPRGKLSVFTSKQNRQMGVSGKNTKSTISGDMCSLEKNDGSDTAVGA 1317
 Db 1 MSVQSEPRGKLSVFTSKQNRQMGVSGKNTKSTISGDMCSLEKNDGSDTAVGT 60
 QY 1318 LQTSKKRRSISGAKVAIVLSKRSASQSLQTEGGKKLSTVORSTETGLAVEMRN 1377
 Db 61 LQTSKKRRSISGAKVAIVLSKRSASQSLQTEAGGKKLSTVORSTETGLAVEMRN 120
 QY 1378 WMTQASRESTDGSMNSYSSEGNLIFFGVRLASDQPSDFLDGLGPAQLVGRQTLATPAM 1437
 Db 121 WMTQASRESTDGSMNSYSSEGNLIFFGVRLASDQPSDFLDGLGPAQLVGRQTLATPAM 180
 QY 1438 GDTQVGMKKGQLEVEIIRARGLVKPGSKTLPAPYVKVYLLDNGVC----- 1485
 Db 181 GDTQVGMKKGQLEVEIIRARGLVKPGSKTLPAPYVKVYLLDNGVCIAKKGQLEVEII 240
 QY 1486 -----TAKKTKVARKTLEPLYQQLLSFESSPOGRV 1516
 Db 241 RARGLVKPGSKTLPAPYVKVYLLDNGVXIATKTKVARKTLEPLYQQLLSFESSPOGRV 300
 QY 1517 LQIIVMGDGRMDHKSFMGVAQILLDELSNMVIGWFKLPPPSLVDPTSAPITRRASQ 1576
 Db 301 LQIIVMGDGRMDHKSFMGVAQILLDELSNMVIGWFKLPPPSLVDPTSAPITRRASQ 360
 QY 1577 SSLESSTGPSYRS 1590
 Db 361 SSLESSTGPSYRS 374
 RESULT 4
 AAB34848
 ID AAB34848 standard; Protein; 237 AA.
 XX AC AAB34848;
 XX

XX 26-JAN-2001 (first entry)
 DE Human secreted protein sequence encoded by gene 46 SEQ ID NO:136.
 XX Human; secreted protein; diagnosis; cytostatic; immunosuppressive;
 KW nototropic; neuroprotective; antiviral; antiallergic; hepatotropic;
 KW antidiabetic; antiinflammatory; antitumor; antiparasitic; anticonvulsant;
 KW antibacterial; antifungal; antiparasitic; cardiant; gene therapy;
 KW cancer; immune disorder; cardiovascular disorder; wound healing;
 KW neurological disease; infectious disease; chromosome identification;
 KW chromosome 6.
 XX Homo sapiens.
 OS WO2000058356-A1.
 XX 05-OCT-2000.
 XX 22-MAR-2000; 2000WO-US07535.
 XX 26-MAR-1999; 99US-0126511.
 XX 17-DEC-1999; 99US-0172413.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX Rosen CA, Ruben SM, Komatsoulis G;
 PI WPI; 2000-594639/56.
 XX Fifty nucleic acid molecules encoding human secreted proteins, useful
 PT in the prevention, treatment and diagnosis of cancer, immune disorders,
 PT cardiovascular disorders and neurological diseases -
 XX Disclosure; Page 419-420; 425pp; English.
 XX The polynucleotide sequences given in AAC59966 to AAC60015 encode the
 CC human secreted proteins given in AAB34773 to AAB34822. AAB34823 to
 CC AAB34852 represent human secreted polypeptide sequences and proteins
 CC homologous to them, which are given in the exemplification of the present
 CC invention. Human secreted proteins have activities based on the tissues
 CC and cells the genes are expressed in. Examples of activities include:
 CC cytostatic; immunosuppressive; nototropic; neuroprotective; antiviral;
 CC antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
 CC vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic; and
 CC cardiant. The polynucleotides and polypeptides are useful for
 CC preventing, treating or ameliorating a medical condition in e.g. humans,
 CC mice, rabbits, goats, horses, cats, dogs, chickens or sheep. The
 CC polypeptides can also be used as a food additive or preservative to
 CC increase or decrease storage capabilities. The polynucleotides are
 CC useful for chromosome identification. They are also useful as probes for
 CC diagnosing a disorder related to the female reproductive system,
 CC particularly breast and/or ovary cancer. They are also useful in the gene
 CC therapy of breast and ovarian cancer. Nucleic acids, protein, antibodies,
 CC agonists and antagonists from the present invention are useful in the
 CC diagnosis, treatment and prevention of cancer, immune disorders,
 CC cardiovascular disorders, wound healing, neurological diseases and
 CC infectious disease. AAC59957 to AAC59965 and AAB34772 represents sequence
 CC used in the exemplification of the present invention.
 XX Sequence 237 AA;
 SQ
 Query Match 12.3%; Score 1015.5; DB 21; Length 237;
 Best Local Similarity 82.8%; Pred. No. 4.6e-63;
 Matches 197; Conservative 19; Mismatches 21; Indels 1; Gaps 1;
 QY 1353 EGGGKLLRSTVORSTETGLAVEMRNMTQASRESTDGSMNSYSSEGNLIFFGVRLASD 1412
 Db 1 EGGGKLLRSTVORSTETGLAVEMRNMTQASRESTDGSMNSYSSEGNLIFFGVRLASD 59
 QY 1413 QPSDFLDGLGPAQLVGRQTLATPAMGDIQVGMKKGQLEVEIIRARGLVKPGSKTLP 1472
 Db 60 QPSDFLDGLGPAQLVGRQTLATPAMGDIQVGMKKGQLEVEIIRARGLVKPGSKTLP 119

QY 1473 PYKVVLLNGVCIKKTKVARKTSLPLVQQLSFFESPGQVRLQIIVWGDYGRMDHKS 1532
Db 120 PYKVVLLNGVCIKKTKVARKTSLPLVQQLSFFESPGQVRLQIIVWGDYGRMDHKS 179
QY 1533 FMGVAQILLDELSELSNNVIGWFKLPPSSSLVDPTSLPTTRASQSSLESSTGSPYSRS 1590
Db 180 FMGVAQILLDELSELSNNVIGWFKLPPSSSLVDPTSLPTTRASQSSLESSTGSPYSRS 237

RESULT 5
AAB34847
ID AAB34847 standard; Protein; 237 AA.
AC AAB34847;
DT 26-JAN-2001 (first entry)
DE Gene 46 human secreted protein homologous amino acid sequence #135.
XX Human; secreted protein; diagnosis; cytostatic; immunosuppressive;
KW neurotropic; neuroprotective; antiviral; antiallergic; hepatotropic;
KW antidiabetic; antiinflammatory; antitumor; vulnery; anticonvulsant;
KW antibacterial; antifungal; antiparasitic; cardiant; gene therapy;
KW cancer; immune disorder; cardiovascular disorder; wound healing;
KW neurological disease; infectious disease; chromosome identification.

OS Rattus norvegicus.
XX W0200058356-A1.
XX 05-OCT-2000.
XX 22-MAR-2000; 2000WO-US07535.
XX 26-MAR-1999; 99US-0126511.
XX 17-DEC-1999; 99US-0172413.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Ruben SM, Komatsoulis G;
XX WPI; 2000-594639/56.
XX Fifty nucleic acid molecules encoding human secreted proteins, useful
XX in the prevention, treatment and diagnosis of cancer, immune disorders,
XX cardiovascular disorders and neurological diseases -
XX Disclosure; Page 418-419; 425pp; English.

XX The polynucleotide sequences given in AAC59966 to AAC60015 encode the
XX human secreted proteins given in AAB34773 to AAB34822. AAB34823 to
XX AAB34852 represent human secreted polypeptide sequences and proteins
XX homologous to them, which are given in the exemplification of the present
XX invention. Human secreted proteins have activities based on the tissues
XX and cells the genes are expressed in. Examples of activities include:
XX cytostatic; immunosuppressive; neurotropic; neuroprotective; antiviral;
XX antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
XX vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic; and
XX cardiant. The polynucleotides and polypeptides are useful for
XX preventing, treating or ameliorating a medical condition in e.g. humans,
XX mice, rabbits, goats, horses, cats, dogs, chickens or sheep. The
XX polypeptides can also be used as a food additive or preservative to
XX increase or decrease storage capabilities. The polynucleotides are
XX useful for chromosome identification. They are also useful as probes for
XX diagnosing a disorder related to the female reproductive system,
XX particularly breast and/or ovary cancer. They are also useful in the gene
XX therapy of breast and ovarian cancer. Nucleic acids, protein, antibodies,
XX agonists and antagonists from the present invention are useful in the
XX diagnosis, treatment and prevention of cancer, immune disorders,
XX cardiovascular disorders, wound healing, neurological diseases and
XX infectious disease. AAC59957 to AAC59965 and AAB34772 represents sequence
XX used in the exemplification of the present invention.

XX SQ Sequence 237 AA;
Query Match 12.2%; Score 1010.5; DB 21; Length 237;
Best Local Similarity 82.4%; Pred. No. 1e-62;
Matches 196; Conservative 20; Mismatches 21; Indels 1; Gaps 1;
QY 1353 EGGKKLRSVQSTSTETGLAVENNMWTRQASRSTDSNNYSYSSGNLIPPGVRLASDS 1412
Db 1 EGGKKLRSVQSTSTETGLAVENNMWTRQASRSTDSNNYSYSSGNLIPPGVRLASDS 59
QY 1413 QPSDFLDGLGPAQLVGRQTLATPAMGDIQVGMDDKKGQLEVEIRARGLVVRGSKTLPA 1472
Db 60 QPSDFLDGLGPAQLVGRQTLATPAMGDIQVGMDDKKGQLEVEIRARGLVVRGSKSTPA 119
QY 1473 PYKVVLLNGVCIKKTKVARKTSLPLVQQLSFFESPGQVRLQIIVWGDYGRMDHKS 1532
Db 120 PYKVVLLNGVCIKKTKVARKTSLPLVQQLSFFESPGQVRLQIIVWGDYGRMDHKS 179
QY 1533 FMGVAQILLDELSELSNNVIGWFKLPPSSSLVDPTSLPTTRASQSSLESSTGSPYSRS 1590
Db 180 FMGVAQILLDELSELSNNVIGWFKLPPSSSLVDPTSLPTTRASQSSLESSTGSPYSRS 237

RESULT 6
AAB63174
ID AAB63174 standard; Protein; 2464 AA.

XX AC AAB63174;
XX 26-MAR-2002 (first entry)
XX Drosophila melanogaster polypeptide SEQ ID NO 16314.
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX Drosophila melanogaster.
XX W0200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US09231.
XX 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX N-PSDB; ABL07277.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX Disclosure; SEQ ID NO 16314; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABBS7737-ABB72072) and the encoded proteins
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

SQ	Sequence	2464 AA;	
	Query Match	11.9%; Score 989; DB 22; Length 2464;	
	Best Local Similarity	19.8%; Pred. No. 1.1e-59;	
	Matches	503; Conservative 201; Mismatches 507; Indels 1330; Gaps 73;	
QY	272	DKRSPSVRDO-NRRIYQSEKEDYSQVPSDGTWFRSPSYADRSOREPOFYEPCH 330	QY 971 QSKKISDSVSDYDCEDGVG-----VVSDFYHNGRDLOSSTLSVPSQVMS--SNHC 1020
DB	20	DLKCPKTSQSLNHPKQNPBEQTPSSVYSHTQSTSTTPTGRLQVKTSLDP--WROQS 77	DB 949 QCRKEVAGVARDYRTVSGIGQSYHNOASATGYTRGGGNGVGAIGPCMSLSQSHSA 1008
QY	331	LNVRDSNRG-----HRHSKEYIVDDVEDVESDEYERORREBEYQARY 374	QY 1021 SPGSPHVRDVLGRT-----RWSPS-----APPOR-----NYEQ 1051
DB	78	INERDLSSGGGGVGGADPELHSHATRLVASADQ-----RDRVELQJHRE----- 127	DB 1009 AFDGTHSGVGAASGGPAYRSTSPRGSLSPPDDRYTDYFVLPVHGSSPRAPSVQ 1068
QY	375	SDPNLARYV-KPOPYEQRIHARYSRAHRRHS-----DVSLANAELEDKSLILMD 429	QY 1052 G-----1052
DB	128	RLQYPLTRTARLDVQR--NSLNRSHRQFSGGAGVGPQA-----169	DB 1069 GPGVSSAAASASQORFQSRSATATPGSPKQRQLPOVPOTSRGAMLRDLGQDFDERL 1128
QY	430	RPSRQSVSERRAAMENQSYSMERTREACQSSYPQRTSNHSPPTPRESPPLDRPDM- 488	QY 1083 -----HRCI-RATG-----HYNTISRMDRH-----RYMDDH-YSSDRD- 1083
DB	170	GGGGAAGAYRAPPL--AAAGGGSSSSTTSS-----SYTQSPQIV 211	DB 1129 ASGGRFGRHRTQPHQATYRSTGSGWERYTGLSDLSMDARMFRHSLSPKDFM 1188
QY	489	RRADSLRQHLDPSAVKTKREKMETLRNDSLSQSESVRPPPPRPHK 540	QY 1084 -----1084
DB	212	IGPGCYDRSGSMASW-----ASAAAATAADAQQAEN-----246	DB 1189 GFGSDMESVVSVTSSAFSTQSERPTSRGLSFPRNENLFGVNSFLGAGCEPITCL 1248
QY	541	SKKGGKORQV-----SLSSSEELASTPEYTSDDVLESESVSEKDSQKQKT-- 591	QY 1084 -----RDCE-----AADRP-----1093
DB	247	SORGGLGATRGCGSSVSSTEAYWDEPATSSSSAAQDSRRPTE-----RRIKTVR 301	DB 1249 RLHSEPGCRANTI EVDVDCDYLPGGAQQLVLEOLEQLAALAAADSPISAGIQP 1308
QY	592	SEQVLSDS-----NTRSERQKRMYYGHSLEE-----DLEWS----- 625	QY 1094 -----1094
DB	302	FADHEASLASSSLPVTAIPLSTAVDVGAATILSGKSLDVGAVGGADWSWEARQ 361	DB 1309 PHQVLVDANSQQLVEQFYVEPTMAEWMELDPLHPSVYFSPSPHLPQPPDIDID 1368
QY	626	EPOIKDSGVDTCSSTLINEHSHSD--KHPVTWQPKQKQDRLIGRIILANKRLNDGSV 680	QY 1094 --YHRSRTEORPPLERTTT-----RSRSRSDPTNLMSM-----PSLWTR 1134
DB	362	GSQDSATKSGIDTSTFTSSSDNRGDKPNVNMQTSADNTRLIGHMILRY--416	DB 1369 TIYPFRSNGLPKPVQATAAAYPLPLPSFQPKITITITNFRSSSPAGVHPHNDHT 1428
QY	681	PRDSGAMGLKVVGK-----MTESGRICAFITKVKKGLSLATVGHLP 724	QY 1135 SAPPSPAL-----SRSHPRTSQVOT-----1154
DB	417	DGEDILGLKVNQGPPLATGSGVSGAGTCAAGACGPGCAIVEKVRGVSADLEGRIRP 474	DB 1429 STTSSLVGVYRDHQDKSCSHCQNGSQPVVLSHTHTLHTAAGLAIGAQQDPCCLADAP 1488
QY	725	QDELEWNGELLOGATFEETVNIILSKPEPOVELVUSRPIG-----766	QY 1155 -----SPSS-----TGTGRCR- 1167
DB	475	GBILEWNGRLANKSADEYDIIDSRLDAQVELIVSRPIGSGGGGSSANVSPISGA 534	DB 1489 HPQHPMGVGEPPSSDPAMCGECVDQHFGLGLTGPQLNLGVVPTVYHVTFTFNAHRRAG 1548
QY	767	DIPRIDST--HAQLESSSS-----FESQK-----MDRPSISV 798	QY 1168 --QLPQLP-----KGTLESAMDIERNRQ 1191
DB	535	SGGSANSVPARRSSANFPHSGGLASSMAGSAVSSGGRYLQKAPAVEATIEHDKPSVLI 594	DB 1549 QWATFAPPTQSVLRLSRLSEDALVAAPVISEAKAQTSLIKKPLERRLFHGQSSS 1608
QY	799	TSPMSFGMLRDV-----810	QY 1192 MKL-----NKYQVA-----GSDPRLQDY 1211
DB	595	TSPGSPDIHTSVSGPGSVSGILQRGGVGTQRLGPHSHSHSHSGSGSGSSTSV 654	DB 1609 LDYDDLHTKSGRRIRIYEDVMPSSGDYFPHSNFYSSYSTAAMSRYSSETMIROSY 1668
QY	811	-----810	QY 1212 -----HSKY-----RSGM-----1219
DB	655	HAPASAGFVHASASAGSAPGLGHTTGAHHHHPHPHYHHPHQPHPHQHQPAA 714	DB 1669 MGANVDANHLSSHFLVTDDKIVITVDSVDGWTGRGVAPELFRGAHRQRGLADPRNM 1728
QY	811	-----PQFLSGQLSILKMFQKQGLIIVILGAKDLPSPREDGR 848	QY 1220 -----DPRGAD-----1226
DB	715	HLQCHGVGGIGMGSGGTTQPIPIEGRLOKLQGLQDNTQLIIVLVCATGLSLRQSGA 774	DB 1729 SLDLQTSQKLYGPAAVYLKERRNLPHPPSAQNAHNSFMEPGSGMDGLAAAGVQ 1788
QY	849	PRNPYKIYLPDRDKNGKRTKTKLEPKNQNTFYSPVHRRFRERMLBITLWQA 908	QY 1227 -----TVSTKSDSDSDVSVAVSRTSS-----1248
DB	775	GRNPYAKVFLPDRSHKSKRTKTVGTCTCEPRWGTQFVYSGLRCDLNGELLEVILWDY- 833	DB 1789 GNGWGGGAHNTTSTHNSHNSQASSVSQSHNQQLQQQQQLQQSVRQKKEKETTAL 1848
QY	909	RYVEESEFEIILETALLDEHMYKLQTH-----DVSSLPL-----949	QY 1249 -----ASRPSSTSYMSVQSERPR-----1266
DB	834	VRYGANDFIGEVVIDLAHHILDDEAWYQLQPHQDTSYLLRDEGSDVGLIILPTDHL 892	DB 1849 SGSMASEQLTKSSSGSGGATSAAGHLVVGGTNNANSSSSANFPANPPPPQQOOLHG 1908
QY	950	-----PRPSPYL-----PRQLHGESPTREL 970	QY 1267 -----GNEKISVETSKMONQWYSGKN-----1289
DB	893	PPSTMFLSDSTTSDCIDGTFGASISSMGSSASPPPLELDELNERRSRDMS-948	DB 1909 SPNVLNNKRAPSTTPOQNTNTINDANMLTDATQSSQQOOLQOQPPNTCTNQNTPT 1968
			QY 1290 -LTKSTSSG-----1298
			DB 1969 TLTTTTSSNATNAATTTTANATNATTTTATNPTTTTTTNTNTPNATTTTTTTTANADAD 2028

PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226686.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.

PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 20-OCT-2000; 2000US-0242221.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250391.
 PR 01-DEC-2000; 2000US-0251160.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 (HUMA-) HUMAN GENOME SCI INC.
 Rosen CA, Barash SC, Ruben SM;
 WPI; 2001-541565/60.
 N-PSDB; ABA11415.
 Nucleic acids encoding 3224 human nervous system antigen polypeptides,
 useful for preventing, diagnosing and/or treating nervous system
 cancers and metastases -
 Claim 11; SEQ ID NO 3746; 1701pp + Sequence Listing; English.

XX
 XX

CC The invention relates to novel genes (AB11004-ABA21534) and proteins (AB14678-AB18001) useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the adrenal gland, bone, marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemia; (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections.

CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 158 AA;

Query Match 6.9%; Score 569.5; DB 22; Length 158;
Best Local Similarity 41.2%; Pred. No. 5.3e-32;
Matches 133; Conservative 3; Mismatches 12; Indels 175; Gaps 3;

QY 1068 MDRHRVMDHYSDRDCEAADROPYHRSRSTQRLPRTTTRSSSTPTNLRSM 1127
DB 1 MDRHRVMDHYSDRDCEAADROPYHRSRSTQRLPRTTTRSSSTPTNLRSM 60

QY 1128 PSLMTGRSAPPSALSRSHPTSGVTSPTGTGRRGRLPOLPKGLRSAMDIER 1187
DB 61 PSMTGRSAPPSALSRSHPTSGVTSPTGTGRRGRLPOLPKGLRKA----- 115

QY 1188 RNRQMLNKYQVAGSDPRLEQDYHYSKRSQWDPHFGADTVTSKSDSDVSVAVSRTS 1247
DB 116 ----- 115

QY 1248 SASRPSSTYSVQSERPRGNRKISVTSKQNRQMVSGNLTSTISGDWCSLEKND 1307
DB 116 ----- 115

QY 1308 GSQSDTAVGALGTGKCRSSIGAKMVAIVGLSKRSASQLSTEGGKKLRSTVORST 1367
DB 116 -----GGKKLRSTVORST 128

QY 1368 ETGLAV--EMRNMWTRQASREST 1388
DB 129 ETGLAVGNEELXW-TRQKPESS 150

RESULT 9
AAB34850

ID AAB34850 standard; Protein; 128 AA.

AC AAB34850;

DT 26-JAN-2001 (first entry)

XX Human secreted protein sequence encoded by gene 46 SRQ ID NO:138.

XX Human; secreted protein; diagnosis; cytostatic; immunosuppressive; neurotropic; neuroprotective; antiviral; anti-allergic; hepatotropic; antidiabetic; antiinflammatory; antitumor; vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic; cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder; wound healing; neurological disease; infectious disease; chromosome identification; chromosome 6.

XX Homo sapiens.

XX WO200058356-A1.

XX 05-OCT-2000.

PD

XX 22-MAR-2000; 2000WO-US07535.
PF
XX 26-MAR-1999; 99US-0126511.
PR 17-DEC-1999; 99US-0172413.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Ruben SM, Komatsculis G;
PI WPI; 2000-594639/56.
XX Fifty nucleic acid molecules encoding human secreted proteins, useful in the prevention, treatment and diagnosis of cancer, immune disorders, cardiovascular disorders and neurological diseases -
XX Disclosure; Page 421; 425pp; English.
XX The polynucleotide sequences given in AAC59966 to AAC60015 encode the human secreted proteins given in AAB34773 to AAB34822. AAB34823 to AAB34852 represent human secreted polypeptide sequences and proteins homologous to them, which are given in the exemplification of the present invention. Human secreted proteins have activities based on the tissues and cells the genes are expressed in. Examples of activities include: cytostatic; immunosuppressive; neurotropic; neuroprotective; antiviral; anti-allergic; hepatotropic; antidiabetic; antiinflammatory; antitumor; vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic; and cardiant. The polynucleotides and polypeptides are useful for preventing, treating or ameliorating a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. The polypeptides can also be used as a food additive or preservative to increase or decrease storage capabilities. The polynucleotides are useful for chromosome identification. They are also useful as probes for diagnosing a disorder related to the female reproductive system, particularly breast and/or ovary cancer. They are also useful in the gene therapy of breast and ovarian cancer. Nucleic acids, protein antibodies, agonists and antagonists from the present invention are useful in the diagnosis, treatment and prevention of cancer, immune disorders and cardiovascular disorders, wound healing, neurological diseases and infectious disease. AAC59957 to AAC59965 and AAB34772 represents sequence used in the exemplification of the present invention.

XX Sequence 128 AA;

Query Match 6.8%; Score 566; DB 21; Length 128;
Best Local Similarity 82.0%; Pred. No. 6.8e-32;
Matches 105; Conservative 13; Mismatches 10; Indels 0; Gaps 0;

QY 1446 DKKGQLEVEILIRAGLVVKGSKTLPAPYKVVLLDNGVCIKKTKVAKTLEPLVQOL 1505
DB 1 DKKGQLEVEIVIRARSUTQKPGSKSTPAPYKVVLLDNGVCIKKTKVAKTLEPLVQOL 60

QY 1506 LSPFESPOQRVLQIVWGYGRMDHKSFMGVAQILLDELRLNNVIGWFKLPPSSLVDP 1565
DB 61 LVPDESPOQRVLQIVWGYGRMDHKSFMGVAQILLDELRLNNVIGWFKLPPSSLVDP 120

QY 1566 TSAPLTER 1573

DB 121 TLTPLTR 128

RESULT 10

ABP63532

ID ABP63532 standard; Protein; 92 AA.

XX ABP63532;

XX 28-OCT-2002 (first entry)

DT KIAA 0751 PDZ domain related amino acid sequence #2.

XX KIAA 0751 PDZ domain related amino acid sequence #2.

XX Molecular interaction; haematopoietic cell; immune response; T cell;
XX PDZ domain; B cell; endothelial cell; PDZ protein; PSD95; PDZ ligand;

allergies, aplastic anaemia, burns, wounds, bone and cartilage damage, nocturnal haemoglobinuria, anti-inflammatory disease; to enhance coagulation; to inhibit thrombosis; and as a contraceptive.

Sequence 126 AA;

Query Match 5.1%; Score 420; DB 21; Length 126;
Best Local Similarity 65.3%; Pred. No. 1.2e-21;
Matches 8; Conservative 19; Mismatches 24; Indels 0; Gaps 0;

1358 KLRSTVQRSTETGLAVERNWMTQASRSTGSMNSYSGEMLIFPGVRLADSPDSDF 1417
1 RVKGAIGRSTETGLAVERNWMTQASRSTGSMNSYSGEMLIFPGVRLADSPDSDF 60

1418 LDGLGPAQLVGRITLTPAMGDTQVGMOKKQGLVEIIPARGLVNPGSKTLPAPYKRV 1477
61 LGSMGPAQFVGRITLTPAMGDTQVGMOKKQGLVEIIPARGLVNPGSKTLPAPYKRV 1477

1478 YLKD 1481
121 YLKE 124

RESULT 12
ABP63179
ID ABP63179 standard; Peptide; 82 AA.
XX AC ABP63179;
XX DT 28-OCT-2002 (first entry)
XX DE KIAA 0751 PDZ domain amino acid sequence.
XX KW Molecular interaction; haematopoietic cell; immune response; T cell;
XX KW PDZ domain; B cell; endothelial cell; PDZ protein; PSD95; PDZ ligand;
XX KW Drosophila large disc protein; Zonula Occludin 1 protein; PL protein;
XX KW immunosuppressive; anti-inflammatory; anti-allergic; anti-atherosclerotic;
XX KW antitumor; antipsoriatic; dermatological; antiasthmatic; cytostatic;
XX KW antimicrobial; vasotropic; inflammatory immune response; inflammation;
XX KW humoral immune response; autoimmune disease; allergy; ulcerative colitis;
XX KW inflammatory bowel disease; ileitis; enteritis; psoriasis; scleroderma;
XX KW inflammatory dermatosis; respiratory allergic disease; asthma; cancer;
XX KW allergic rhinitis; transplantation rejection; atherosclerosis; ischaemia;
XX KW angiogenesis-dependent disorder; infectious disease.
XX OS Homo sapiens.
XX PN WO200242422-A2.
XX PD 30-MAY-2002.
XX PF 09-NOV-2001; 2001WO-US44138.
XX PR 11-NOV-2000; 2000US-0710059.
XX PR 24-NOV-2000; 2000US-0721915.
XX PR 24-NOV-2000; 2000US-0722069.
XX PR 28-NOV-2000; 2000US-0724553.
XX PA (ARBO-) ARBOR VITA CORP.
XX PI Lu P, Rabinowitz JD, Schweizer J;
XX DR WPI; 2002-608221/65.
XX PT Modulating the biological function of an endothelial cell or
XX PT haematopoietic cell e.g., a T-cell or B-cell comprises introducing into
XX PT the cell, an agent that inhibits binding of a PDZ protein and a PDZ
XX PT ligand protein in the cell -
XX PS Disclosure; Page 47; 207pp; English.
XX CC The present invention describes a method (M1) for modulating a biological
XX CC function of an endothelial cell or haematopoietic cell. M1 comprises

introducing into the cell, an agent that inhibits binding of a PDZ (PSD95, Drosophila large disc protein, and Zonula Occludin 1 protein) protein and a PDZ ligand (PL) protein in the cell, and so modulates the biological function. Also described is a method (M2) for determining whether a test compound is an inhibitor of binding between a PDZ protein and a PL protein. M1 is used for modulating a biological function of an endothelial cell or haematopoietic cell e.g., T-cell or B-cell, by an inflammatory or humoral immune response, or an autoimmune disease. An inhibitor (I) is useful for treating a disease characterised by leukocyte activation, where the disease is characterised by an inflammatory or humoral immune response, e.g., an autoimmune disease. The compounds e.g., PL-PDZ interaction inhibitors are useful for treating (ameliorating symptoms of) a variety of diseases and conditions characterised by inflammatory and humoral immune responses e.g., inflammation, allergy, inflammatory bowel diseases, ulcerative colitis, ileitis and enteritis, psoriasis and inflammatory dermatoses, scleroderma, respiratory allergic diseases such as asthma, allergic rhinitis, transplantation rejection (cardiac, kidney, lung, liver, small bowel, cornea, pancreas, cadaver, autologous, bone marrow, xenotransplantation), atherosclerosis, cancers, angiogenesis-dependent disorders, infectious diseases and ischaemia. CC ABQ96620 to ABQ96732 and ABP63153 to ABP63578 represent sequences used CC in the exemplification of the present invention.

Sequence 82 AA;

Query Match 5.0%; Score 417; DB 23; Length 82;
Best Local Similarity 100.0%; Pred. No. 1e-21;
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

682 RDSGAMGLKVVGGKMTESGRICAFITKVKGLADTVGHLRPGDVLWNGRLQGATP 741
1 RDSGAMGLKVVGGKMTESGRICAFITKVKGLADTVGHLRPGDVLWNGRLQGATP 60

742 KEVYNILLESKEPQVELVVS 763

61 KEVYNILLESKEPQVELVVS 82

RESULT 13

AAB55848

ID AAB55848 standard; Peptide; 86 AA.

XX AC AAB55848;

XX DT 07-MAR-2001 (first entry)

XX DE PDZ encoded domain #18.

XX KW Hematopoietic cell; PDZ; PL; autoimmune disease; inflammation;
XX KW allergy; asthma; multiple sclerosis; cancer; infection.
XX OS Synthetic.

XX PN WO2000069896-A2.

XX PD 23-NOV-2000.

XX PF 12-MAY-2000; 2000WO-US13161.

XX PR 14-MAY-1999; 99US-0134114.

XX PR 14-MAY-1999; 99US-0134117.

XX PR 14-MAY-1999; 99US-0134118.

XX PR 21-OCT-1999; 99US-0160860.

XX PR 23-OCT-1999; 99US-0162498.

XX PR 13-DEC-1999; 99US-0170453.

XX PR 14-JAN-2000; 2000US-0176195.

XX PR 14-FEB-2000; 2000US-0182296.

XX PR 11-APR-2000; 2000US-0196460.

XX PR 11-APR-2000; 2000US-0196527.

XX PA (ARBO-) ARBOR VITA CORP.

XX PI Lu PS;

XX WFI; 2001-080245/09.
 XX
 XX Modulating a biological function of an endothelial cell or
 PT hematopoietic cell, useful for treating autoimmune diseases and
 PT infectious diseases, by administering an antagonist that inhibits
 PT binding between a PDZ protein and a PL protein -
 XX
 XX Disclosure; Page 28-43; 141pp; English.
 XX
 XX The present invention relates to a new method for modulating a
 CC biological function of an endothelial cell or hematopoietic cell. The
 CC method involves introducing into a cell, an antagonist that inhibits
 CC binding between a PDZ protein and a PL protein. The inhibitor is used
 CC to treat a disease mediated by hematopoietic cells, e.g. autoimmune
 CC disease. It may also be used to prevent transplantation rejection of
 CC a solid organ transplant. The method may also be used in the treatment
 CC of inflammation, allergy, inflammatory bowel diseases, ulcerative
 CC colitis, ileitis, psoriasis, asthma, atopic dermatitis, autoimmune
 CC diseases (e.g. rheumatoid arthritis, multiple sclerosis,
 CC insulin-dependent diabetes, Hashimoto thyroiditis, osteoarthritis,
 CC graft rejection, transplantation rejection), atherosclerosis, cancers,
 CC infectious diseases, ischemia, vasculitis and Crohn's disease.
 XX
 XX Sequence 86 AA;
 SQ
 Query Match 5.0%; Score 417; DB 22; Length 86;
 Best Local Similarity 100.0%; Pred. No. 1.1e-21;
 Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 682 RDGAMGLKLVVGGKMTESGRLCAPITTKVKGSLADTVGHLRPGDVEVLEWNGRLQGATF 741
 DB 1 RDGAMGLKLVVGGKMTESGRLCAPITTKVKGSLADTVGHLRPGDVEVLEWNGRLQGATF 60
 QY 742 EEVYNILLESKPQVELVWSR 763
 DB 61 EEVYNILLESKPQVELVWSR 82
 RESULT 14
 AAB57639
 ID AAB57639 standard; Protein; 86 AA.
 XX
 XX AAB57639;
 XX
 XX 12-MAR-2001 (first entry)
 DE PDZ domain.
 XX
 XX Endothelial cell; hematopoietic cell; PDZ domain protein;
 KW PL domain protein; leukocyte activation; synapse formation;
 KW transmembrane neurotransmitter receptor; autoimmune disease;
 KW transplantation rejection; inflammation; allergy;
 KW inflammatory bowel disease; ulcerative colitis; ileitis; psoriasis;
 KW asthma; atopic dermatitis; atherosclerosis; cancer; infectious disease;
 KW ischaemia; vasculitis; Crohn's disease.
 OS Homo sapiens.
 XX
 XX WO200069897-A2.
 PN
 XX 23-NOV-2000.
 PD
 XX 12-MAY-2000; 2000WO-US13166.
 XX
 XX 14-MAY-1999; 99US-0134114.
 PR 14-MAY-1999; 99US-0134117.
 PR 14-MAY-1999; 99US-0134118.
 PR 21-OCT-1999; 99US-0160860.
 PR 29-OCT-1999; 99US-0162498.
 PR 13-DEC-1999; 99US-0170453.
 PR 14-JAN-2000; 2000US-0176195.
 PR 14-FEB-2000; 2000US-0182296.
 PR

PR 11-APR-2000; 2000US-0196460.
 PR 11-APR-2000; 2000US-0196527.
 XX (ARBO-) ARBOR VITA CORP.
 XX
 XX Lu PS;
 PT
 DR WFI; 2001-025003/03.
 XX
 XX New inhibitors of binding of a PDZ protein and PL protein for
 PT inhibiting T cell-mediated responses by hematopoietic cells, or for
 PT treating diseases characterized by inflammatory and humoral immune
 PT responses, e.g. inflammation, cancer -
 XX
 XX Disclosure; Page 36; 139pp; English.
 XX
 XX The present invention relates to a method for modulating a biological
 CC function of an endothelial cell or hematopoietic cell, comprises
 CC introducing into a cell an antagonist that inhibits binding between a
 CC PDZ domain protein and a PL domain protein to result in inhibition of
 CC leukocyte activation. The present sequence is a PDZ domain. PDZ domains
 CC of proteins are named after three prototypical proteins: PSD95,
 CC Drosophila large disc protein and Zonula Occludin 1 protein. PDZ domain
 CC proteins are involved in synapse formation by organising transmembrane
 CC neurotransmitter receptors through intracellular interactions. The
 CC inhibitors identified by the present invention can be used to treat a
 CC disease mediated by hematopoietic cells, e.g. autoimmune disease,
 CC inflammation, allergy (e.g. drug allergies), inflammatory bowel diseases,
 CC ulcerative colitis, ileitis, psoriasis, respiratory allergic diseases
 CC (e.g. asthma), atopic dermatitis, autoimmune diseases (e.g. rheumatoid
 CC arthritis, multiple sclerosis, insulin-dependent diabetes, Hashimoto
 CC thyroiditis, osteoarthritis), atherosclerosis, cancers, infectious
 CC diseases (e.g. viral infection), ischaemia, vasculitis and Crohn's
 CC disease. The inhibitors can also be used to prevent transplantation
 CC rejection of a solid organ transplant.
 XX
 XX Sequence 86 AA;
 SQ
 Query Match 5.0%; Score 417; DB 22; Length 86;
 Best Local Similarity 100.0%; Pred. No. 1.1e-21;
 Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 682 RDGAMGLKLVVGGKMTESGRLCAPITTKVKGSLADTVGHLRPGDVEVLEWNGRLQGATF 741
 DB 1 RDGAMGLKLVVGGKMTESGRLCAPITTKVKGSLADTVGHLRPGDVEVLEWNGRLQGATF 60
 QY 742 EEVYNILLESKPQVELVWSR 763
 DB 61 EEVYNILLESKPQVELVWSR 82
 RESULT 15
 AAB58049
 ID AAB58049 standard; Protein; 86 AA.
 XX
 XX AAB58049;
 XX
 XX 12-MAR-2001 (first entry)
 DE PDZ domain.
 XX
 XX Endothelial cell; hematopoietic cell; PDZ domain protein;
 KW PL domain protein; leukocyte activation; synapse formation;
 KW transmembrane neurotransmitter receptor; autoimmune disease;
 KW transplantation rejection; inflammation; allergy;
 KW inflammatory bowel disease; ulcerative colitis; ileitis; psoriasis;
 KW asthma; atopic dermatitis; atherosclerosis; cancer; infectious disease;
 KW ischaemia; vasculitis; Crohn's disease.
 OS Homo sapiens.
 XX
 XX WO200069898-A2.
 PN
 XX 23-NOV-2000.
 PD
 XX 12-MAY-2000; 2000WO-US13166.
 XX
 XX 14-MAY-1999; 99US-0134114.
 PR 14-MAY-1999; 99US-0134117.
 PR 14-MAY-1999; 99US-0134118.
 PR 21-OCT-1999; 99US-0160860.
 PR 29-OCT-1999; 99US-0162498.
 PR 13-DEC-1999; 99US-0170453.
 PR 14-JAN-2000; 2000US-0176195.
 PR 14-FEB-2000; 2000US-0182296.
 PR

PD 23-NOV-2000.
XX
XX
XX 12-MAY-2000; 2000WO-US13205.
XX
XX 14-MAY-1999; 99US-0134114.
XX 14-MAY-1999; 99US-0134117.
XX 14-MAY-1999; 99US-0134118.
XX 21-OCT-1999; 99US-0160860.
XX 29-OCT-1999; 99US-0162498.
XX 13-DEC-1999; 99US-0170453.
XX 14-JAN-2000; 2000US-0176195.
XX 14-FEB-2000; 2000US-0182296.
XX 11-APR-2000; 2000US-0196460.
XX 11-APR-2000; 2000US-0196527.
XX
XX (ARBO-) ARBOR VITA CORP.
XX
XX Lu PS;
XX
XX WPI; 2001-061214/07.
XX
XX Modulating a biological function of a hematopoietic cell for treating
XX an allergic response, or diseases mediated by immune system cells,
XX comprises introducing into the cell a PDZ-PL interaction enhancer or
XX inhibitor -
XX
XX Disclosure; Page 38; 143pp; English.
XX
XX The present invention relates to a method for modulating a biological
XX function of an endothelial cell or hematopoietic cell, comprises
XX introducing into a cell an antagonist that inhibits binding between a
XX PDZ domain protein and a PL domain protein to result in inhibition of
XX leukocyte activation. The present sequence is a PDZ domain. PDZ domains
XX of proteins are named after three prototypical proteins: PSD95,
XX Drosophila large disc protein and Zonula Occludin 1 protein. PDZ domain
XX proteins are involved in synapse formation by organising transmembrane
XX neurotransmitter receptors through intracellular interactions. The
XX inhibitors identified by the present invention can be used to treat a
XX disease mediated by hematopoietic cells, e.g. autoimmune disease,
XX inflammation, allergy (e.g. drug allergies), inflammatory bowel diseases,
XX ulcerative colitis, ileitis, psoriasis, respiratory allergic diseases
XX (e.g. asthma), atopic dermatitis, autoimmune diseases (e.g. rheumatoid
XX arthritis, multiple sclerosis, insulin-dependent diabetes, Hashimoto
XX thyroiditis, osteoarthritis), atherosclerosis, cancers, infectious
XX diseases (e.g. viral infection), ischaemia, vasculitis and Crohn's
XX disease. The inhibitors can also be used to prevent transplantation
XX rejection of a solid organ transplant.
XX
XX Sequence 86 AA;
XX
XX Query Match 5.0%; Score 417; DB 22; Length 86;
XX Best Local Similarity 100.0%; Pred. No. 1.1e-21;
XX Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 682 RDGAMGLGVGGKMTESGRLCAFTKVKKGLADTVGHLPDGVLEWNGRLQCATF 741
XX Db 1 RDGAMGLGVGGKMTESGRLCAFTKVKKGLADTVGHLPDGVLEWNGRLQCATF 60
XX
XX QY 742 EYVNIILSKPEPQVELVYSR 763
XX Db 61 EYVNIILSKPEPQVELVYSR 82

Search completed: December 12, 2003, 07:13:57
Job time : 84 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 11, 2003, 15:16:19 ; Search time 17412 Seconds
(without alignments)
11700.548 Million cell updates/sec

Title: US-09-617-099B-2
Perfect score: 4980
Sequence: 1 gcttcctaggggtggtcgg.....acattgtgcctagcaagg 4980

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0
2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: sb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inv.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pln.*
- 35: em.htg.rod.*
- 36: em.htg.man.*
- 37: em.htg.vrt.*
- 38: em.sy.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4980	100.0	✓ 4980	6	AX137063 Sequence
2	4980	100.0	✓ 4980	6	B63816 Protein Rim
3	4608.4	92.5	4800	10	AB021131 Mus muscu
4	3886	78.0	5550	10	AF199331 Rattus no
5	3834	77.0	5592	10	AF199330 Rattus no
6	3776	75.8	5640	10	AF199322 Rattus no
7	3247.8	65.2	5063	10	AF548738 Rattus no
8	3197.6	64.2	5172	10	AF199324 Rattus no
9	2936.4	59.0	4965	10	AF199329 Rattus no
10	2860.4	57.4	5031	10	AF199323 Rattus no
11	2860.4	57.4	5031	10	AF199326 Rattus no
12	2860.4	57.4	5031	10	AF199332 Rattus no
13	2860.4	57.4	5031	10	AF199332 Rattus no
14	2802.4	56.3	5079	10	AF199325 Rattus no
15	2631.2	52.8	4851	10	AF199328 Rattus no
16	2364.8	47.5	3854	9	AB018294 Homo sapi
17	2040.4	41.0	4456	9	BC043144 Homo sapi
18	1204	24.2	5655	10	AF007836 Rattus no
19	1075.8	21.6	5325	9	AB051866 Homo sapi
20	1038.8	20.9	5841	10	AF199333 Rattus no
21	940.2	18.9	1877	9	AF007156 Homo sapi
22	929	18.7	254783	2	AC101935 Mus muscu
23	847.4	17.0	139527	2	AC141382 Rattus no
24	816.2	16.4	5079	9	AY190519 Homo sapi
25	796	16.0	2223	9	AF263306 Homo sapi
26	782.4	15.7	1963	10	AF199335 Rattus no
27	760.6	15.3	2301	9	AF263305 Homo sapi
28	741.6	14.9	139338	9	AC025836 Homo sapi
29	741.6	14.9	152336	9	AF004714 Homo sapi
30	741.6	14.9	162489	9	AP001572 Homo sapi
31	715	14.4	6395	9	AB002338 Human mRN
32	715	14.4	6638	6	BD083751 Nucleic a
33	715	14.4	6638	6	BD097398 Nucleic a
34	715	14.4	6638	9	AB045726 Homo sapi
35	683.6	13.7	2103	9	AF263307 Homo sapi
36	441.2	8.9	526	6	BD109094 EST and e
37	398.2	8.0	1424	9	BC003103 Homo sapi
38	398.2	8.0	7239	9	D87074 Human mRN
39	398.2	8.0	7325	6	AX281663 Sequence
40	397.4	8.0	189034	2	AC14537 Mus muscu
41	379.6	7.6	1605	10	AF199334 Rattus no
42	377	7.6	1551	9	AF263310 Homo sapi
43	377	7.6	1704	9	AF263309 Homo sapi
44	377	7.6	1791	9	AF263308 Homo sapi
45	337.8	6.8	810	10	AF548739 Rattus no

ALIGNMENTS

RESULT 1
AX137063
LOCUS AX137063
DEFINITION Sequence 2 from Patent EP1090986.
ACCESSION AX137063
VERSION AX137063.1 GI:14273409
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE 1
AUTHORS Seino, S., Shibasaki, T. and Ozaki, N.
TITLE Protein rim 2
JOURNAL Patent: EP 1090986-A 2 11-APR-2001;
linear PAT 30-MAY-2001

Db 1981 |||||TAAAGCACTGTGACCTGGAGCCATCCAAAGATGGAGATCGCCTTAATTTGGTCTGATTTT 2040
Qy 2041 |||||ATTAAATAGCGTTTAAAGATGGAGTGTACCTCGAGACTCAGGAGCAATGCTGGGCTT 2100
Db 2041 |||||ATTAAATAGCGTTTAAAGATGGAGTGTACCTCGAGACTCAGGAGCAATGCTGGGCTT 2100
Qy 2101 |||||AAGGTGTAGGAGGAAGATGATGATCAATCAGTGCATCTTGTGCAATTTATACCAAGT 2160
Db 2101 |||||AAGGTGTAGGAGGAAGATGATGATCAATCAGTGCATCTTGTGCAATTTATACCAAGT 2160
Qy 2161 |||||AAAAAGGAAGTTTATGCTGATGATCTAGGACATCTTAGACCAAGTGTAGTCTTTGGA 2220
Db 2161 |||||AAAAAGGAAGTTTATGCTGATGATCTAGGACATCTTAGACCAAGTGTAGTCTTTGGA 2220
Qy 2221 |||||ATGGAATGGAGGCTATTGCAAGAGCCCAATTTGAGGAAGTTTACACATTTATCTAGA 2280
Db 2221 |||||ATGGAATGGAGGCTATTGCAAGAGCCCAATTTGAGGAAGTTTACACATTTATCTAGA 2280
Qy 2281 |||||ATCCAAACCTGAAACCAAGTTGAGCTTGTGTTTCAAGCCCAATTTGGAGATTTCTAG 2340
Db 2281 |||||ATCCAAACCTGAAACCAAGTTGAGCTTGTGTTTCAAGCCCAATTTGGAGATTTCTAG 2340
Qy 2341 |||||AATACCTGTAGCAAGCATGCAAACTGGAATCCAGTTCCTAGCTCATTTGAATCTCAAA 2400
Db 2341 |||||AATACCTGTAGCAAGCATGCAAACTGGAATCCAGTTCCTAGCTCATTTGAATCTCAAA 2400
Qy 2401 |||||AATGGAACCGTCTTCTATATCCGTTTACCTCACCATGAGTCTCCCTGAGGAGTGT 2460
Db 2401 |||||AATGGAACCGTCTTCTATATCCGTTTACCTCACCATGAGTCTCCCTGAGGAGTGT 2460
Qy 2461 |||||CCGCGAGTCTTATCTGGACAGCTTTCAATAAACTATGTTTGAACAGTGTGTCACCA 2520
Db 2461 |||||CCGCGAGTCTTATCTGGACAGCTTTCAATAAACTATGTTTGAACAGTGTGTCACCA 2520
Qy 2521 |||||GTTGATAGTTACAAATTTGGAGCAAGGATCTCCCTCCAGGAGAGTGGAGGCGCAAG 2580
Db 2521 |||||GTTGATAGTTACAAATTTGGAGCAAGGATCTCCCTCCAGGAGAGTGGAGGCGCAAG 2580
Qy 2581 |||||GAATCCTTATGTPAAGATTTACTTCTTCAGATAGAGTGTATAAATAAGAGAGAAC 2640
Db 2581 |||||GAATCCTTATGTPAAGATTTACTTCTTCAGATAGAGTGTATAAATAAGAGAGAAC 2640
Qy 2641 |||||AAAAAGTCAAGAAACTTTGGAACCCAAATGGAACCCAGCTTCAATTTATCTCTGT 2700
Db 2641 |||||AAAAAGTCAAGAAACTTTGGAACCCAAATGGAACCCAGCTTCAATTTATCTCTGT 2700
Qy 2701 |||||CCACCAAGAGATTCGGTGAAAGATCTGGAATTTACCTTTGGGATCAAGCTAGAGT 2760
Db 2701 |||||CCACCAAGAGATTCGGTGAAAGATCTGGAATTTACCTTTGGGATCAAGCTAGAGT 2760
Qy 2761 |||||TCGAGAGAGAGAGAGCAATTTCTTAGAGAGATTTTAAATTTGAAATCGAAACAGCTTTGCT 2820
Db 2761 |||||TCGAGAGAGAGAGAGCAATTTCTTAGAGAGATTTTAAATTTGAAATCGAAACAGCTTTGCT 2820
Qy 2821 |||||AGATGATAGCCGCACTGGTATAGCTGAGACCCATGATGTTCTCTCATTTGCACTCCC 2880
Db 2821 |||||AGATGATAGCCGCACTGGTATAGCTGAGACCCATGATGTTCTCTCATTTGCACTCCC 2880
Qy 2881 |||||TCGCCCTTCCCATATCTGCCCGAGGAGCTCCATGAGAGAGCCCAACCGGAGGCT 2940
Db 2881 |||||TCGCCCTTCCCATATCTGCCCGAGGAGCTCCATGAGAGAGCCCAACCGGAGGCT 2940
Qy 2941 |||||GCAAGGTTCGAGAGATTAAGTGAAGTGTCTGACTAGCACTGGAGAGATGGGCT 3000
Db 2941 |||||GCAAGGTTCGAGAGATTAAGTGAAGTGTCTGACTAGCACTGGAGAGATGGGCT 3000
Qy 3001 |||||GGAGTAGTGTGAGATTCGACAAATGGCCGAGTCTTCAAAGCTTCAAGTCTGCTGCT 3060
Db 3001 |||||GGAGTAGTGTGAGATTCGACAAATGGCCGAGTCTTCAAAGCTTCAAGTCTGCTGCT 3060
Qy 3061 |||||GCCAGACAGTGTATGTCATCAATCAATGCTCAACCATCAGGCTCTCTCATCGATAG 3120

Db 3061 |||||GCCAGACAGTGTATGTCATCAATCAATGCTCAACATCGAGGCTCTCTCATCGATAG 3120
Qy 3121 |||||TGTATTAGAGAGACAGAGTCAATGTCGCTTAGTGCCTCTCTCTCAAGGATTTGGA 3180
Db 3121 |||||TGTATTAGAGAGACAGAGTCAATGTCGCTTAGTGCCTCTCTCTCAAGGATTTGGA 3180
Qy 3181 |||||ACAGGGGACCCGAGGAGAGCTGCTACTGCGCATTTACACACATTTAGCCGCAATGATAG 3240
Db 3181 |||||ACAGGGGACCCGAGGAGAGCTGCTACTGCGCATTTACACACATTTAGCCGCAATGATAG 3240
Qy 3241 |||||ACACCGTGTATGAGTACCACTACTCTTCAGATAGAGACAGGATTTGAGAGAGAGCA 3300
Db 3241 |||||ACACCGTGTATGAGTACCACTACTCTTCAGATAGAGACAGGATTTGAGAGAGAGCA 3300
Qy 3301 |||||TAGACAGCCATATCAAGATCCAGATCAACAGACAAAGCGCTCTCTCTAGAGCGGACAC 3360
Db 3301 |||||TAGACAGCCATATCAAGATCCAGATCAACAGACAAAGCGCTCTCTCTAGAGCGGACAC 3360
Qy 3361 |||||CACCGCTCCAGATCTCTGAAAGCTCTGATACAACTCATGAGTCAATGCTTCAAT 3420
Db 3361 |||||CACCGCTCCAGATCTCTGAAAGCTCTGATACAACTCATGAGTCAATGCTTCAAT 3420
Qy 3421 |||||AATGACTGAGAGATCTGCGCTCTCTCTCACTGCTTATCGAGTCTCACTCTGCTACCG 3480
Db 3421 |||||AATGACTGAGAGATCTGCGCTCTCTCTCACTGCTTATCGAGTCTCACTCTGCTACCG 3480
Qy 3481 |||||GTCTGTCCAGACAAAGCCCATCAAGTACTTCGCGGAAACAGAGAGAGGCGGACAGCTTC 3540
Db 3481 |||||GTCTGTCCAGACAAAGCCCATCAAGTACTTCGCGGAAACAGAGAGAGGCGGACAGCTTC 3540
Qy 3541 |||||ACAGCTTCCACCAAGGAGCAATTTGAGAGAGTGTCTATGATATAGAGAGAGAAATCG 3600
Db 3541 |||||ACAGCTTCCACCAAGGAGCAATTTGAGAGAGTGTCTATGATATAGAGAGAGAAATCG 3600
Qy 3601 |||||CCAAATGAACCTTAAACAAATACAAACAGCTAGCCGCTCAGCCCAAGCTCTGAGAGCA 3660
Db 3601 |||||CCAAATGAACCTTAAACAAATACAAACAGCTAGCCGCTCAGCCCAAGCTCTGAGAGCA 3660
Qy 3661 |||||TTACCAATTCGAAGTATCGCTCAGAGTGGATCCACATAGAGGCGGACAGTCTGTTTCCAC 3720
Db 3661 |||||TTACCAATTCGAAGTATCGCTCAGAGTGGATCCACATAGAGGCGGACAGTCTGTTTCCAC 3720
Qy 3721 |||||TAAATCTCTCGACAGTGTAGTGTATCTGCGTTTCAAGACATGATGCTTCTC 3780
Db 3721 |||||TAAATCTCTCGACAGTGTAGTGTATCTGCGTTTCAAGACATGATGCTTCTC 3780
Qy 3781 |||||TGTTTTTCCAGCACCAAGCTACATGCTCCCTCAATCAGAGCGGCGGAGAGAGAAAGGAA 3840
Db 3781 |||||TGTTTTTCCAGCACCAAGCTACATGCTCCCTCAATCAGAGCGGCGGAGAGAGAAAGGAA 3840
Qy 3841 |||||AATCAGTGTCTTTTACATCCAAATCCAAACAGAGTGGCGCTGTCTGCGGAGAGAACTT 3900
Db 3841 |||||AATCAGTGTCTTTTACATCCAAATCCAAACAGAGTGGCGCTGTCTGCGGAGAGAACTT 3900
Qy 3901 |||||GACCAAAAGCACCAAGTCAAGTGTGAGACATGCTGCTCACTGAGAGAGATGACGAGCCA 3960
Db 3901 |||||GACCAAAAGCACCAAGTCAAGTGTGAGACATGCTGCTCACTGAGAGAGATGACGAGCCA 3960
Qy 3961 |||||GTCCGACACCTGAGTGGCGCCCTGCTGCTACAGTGTGCAAGAGAGCGGATCTAGCAATTGG 4020
Db 3961 |||||GTCCGACACCTGAGTGGCGCCCTGCTGCTACAGTGTGCAAGAGAGCGGATCTAGCAATTGG 4020
Qy 4021 |||||GGCCAAAATGGTAGCTTATTTGTTGCTCTCTCAGGAAAAGTCCAGTGTCTCTCAACTCAG 4080
Db 4021 |||||GGCCAAAATGGTAGCTTATTTGTTGCTCTCTCAGGAAAAGTCCAGTGTCTCTCAACTCAG 4080
Qy 4081 |||||CCAAACCGAGAGAGAGGTTAAAGACTACGAGACTGTCTTACAGAGACACCGAGACCGG 4140
Db 4081 |||||CCAAACCGAGAGAGAGGTTAAAGACTACGAGACTGTCTTACAGAGACACCGAGACCGG 4140
Qy 4141 |||||GCTTACAGTGTGAGATGAGGAACTGATGACCGCCAGCGGAGGATCCACAGATGG 4200
Db 4141 |||||GCTTACAGTGTGAGATGAGGAACTGATGACCGCCAGCGGAGGATCCACAGATGG 4200

Db 3001 ||||| GGGAGTAGTGCAGATTATCGACAAATGGCGCGGATCTTCAAGCTCCAGTGTTCGGT 3060
 QY 3061 ||||| GCCAGAACAGTTCATCATCAATTCATTCATCCATCAGGCTCTCTCATCGAGTAGA 3120
 Db 3061 ||||| GCCAGAACAGTTCATCATCAATTCATTCATCCATCAGGCTCTCTCATCGAGTAGA 3120
 QY 3121 ||||| TGTATAGGAAGGACAAAGTTCATGTGCTGCTAGTGCCTCTCTCTCAAAAGGAATGTGA 3180
 Db 3121 ||||| TGTATAGGAAGGACAAAGTTCATGTGCTGCTAGTGCCTCTCTCTCAAAAGGAATGTGA 3180
 QY 3181 ||||| ACAGGGGACCGAGGGACACGTGCTACTCTGGCCATTAACAACAATTCAGCCGGAATGTAG 3240
 Db 3181 ||||| ACAGGGGACCGAGGGACACGTGCTACTCTGGCCATTAACAACAATTCAGCCGGAATGTAG 3240
 QY 3241 ||||| ACACCGTGTCTGATGATGACCTACTCTTCAGATAGACAGGATTCAGAGCAGCAGA 3300
 Db 3241 ||||| ACACCGTGTCTGATGATGACCTACTCTTCAGATAGACAGGATTCAGAGCAGCAGA 3300
 QY 3301 ||||| TAGACAGCCATATACAGATCCAGATCAACAGAACACGCGCTCTCTAGAGCGACACAC 3360
 Db 3301 ||||| TAGACAGCCATATACAGATCCAGATCAACAGAACACGCGCTCTCTAGAGCGACACAC 3360
 QY 3361 ||||| CACCGCTCCAGATCTCTGAAAGCTCTGATACAAACCTCATGAGTTCGCTTCATT 3420
 Db 3361 ||||| CACCGCTCCAGATCTCTGAAAGCTCTGATACAAACCTCATGAGTTCGCTTCATT 3420
 QY 3421 ||||| AATGACTGGAAGATCTGCGCCCTCTTCACTCTGCTTATCAGGCTCTCACTCTGCTGACCG 3480
 Db 3421 ||||| AATGACTGGAAGATCTGCGCCCTCTTCACTCTGCTTATCAGGCTCTCACTCTGCTGACCG 3480
 QY 3481 ||||| GTCTGTCCAGACAAAGCCATCAAGTACTCCGGGACAGGACGAGGGGCGGACAGCTTCC 3540
 Db 3481 ||||| GTCTGTCCAGACAAAGCCATCAAGTACTCCGGGACAGGACGAGGGGCGGACAGCTTCC 3540
 QY 3541 ||||| ACAGCTTCCACAAAGGGAAACATTTGAGAGAGATGCTATGATATAGAGAGAGAAATCG 3600
 Db 3541 ||||| ACAGCTTCCACAAAGGGAAACATTTGAGAGAGATGCTATGATATAGAGAGAGAAATCG 3600
 QY 3601 ||||| CCAATGAACTTACAAATACAAACAGGTAGCGGATCAGACCCAGACTGGAGCAGA 3660
 Db 3601 ||||| CCAATGAACTTACAAATACAAACAGGTAGCGGATCAGACCCAGACTGGAGCAGA 3660
 QY 3661 ||||| TTACCAATTCGAAGTATCGCTCAGGATGGGATCCACATAGAGGGGAGATCTGTTTCCAC 3720
 Db 3661 ||||| TTACCAATTCGAAGTATCGCTCAGGATGGGATCCACATAGAGGGGAGATCTGTTTCCAC 3720
 QY 3721 ||||| TAAATCCTCGGACAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3780
 Db 3721 ||||| TAAATCCTCGGACAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3780
 QY 3781 ||||| TCGTTTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 3840
 Db 3781 ||||| TCGTTTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 3840
 QY 3841 ||||| AATCAGTGTCTTTACATCCAAATGCAAAACAGACAGATGGCGTGTGCGGGAGAACTT 3900
 Db 3841 ||||| AATCAGTGTCTTTACATCCAAATGCAAAACAGACAGATGGCGTGTGCGGGAGAACTT 3900
 QY 3901 ||||| GACCAAAAGCACAGCATCATGTGGAGACATGTGCTCATCTGGAGAGATGACCGGACCA 3960
 Db 3901 ||||| GACCAAAAGCACAGCATCATGTGGAGACATGTGCTCATCTGGAGAGATGACCGGACCA 3960
 QY 3961 ||||| GTCCGACACTGAGTGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4020
 Db 3961 ||||| GTCCGACACTGAGTGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4020
 QY 4021 ||||| GGCACAAATGCTAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4080
 Db 4021 ||||| GGCACAAATGCTAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4080
 QY 4081 ||||| CCAAAACGAGAGGAGGTAAAGCTTACAGGACATGTTTACAGAGACGACCGGACCGG 4140

4081 CCAAAACGAGAGGAGGTAAAGCTTACGAGACATGTTTACAGAGAGCAGGAGACGG 4140
 4141 GCTAGCAGTGCAGATCAGAGAACTGATGACCCGCGCAGCCGCGGAGATTCACAGATGG 4200
 4141 GCTAGCAGTGCAGATCAGAGAACTGATGACCCGCGCAGCCGCGGAGATTCACAGATGG 4200
 4201 CAGCATGAAACAGCTATAGCTCGGAGGAAATCTGATCTTCCCTGGGCTCGGCTCGGCTC 4260
 4201 CAGCATGAAACAGCTATAGCTCGGAGGAAATCTGATCTTCCCTGGGCTCGGCTCGGCTC 4260
 4261 TGACAGCAGTTCAGTATGATCTTCCCTGGGCTCGGCTCGGCTCGGCTCGGCTCGGCTC 4320
 4261 TGACAGCAGTTCAGTATGATCTTCCCTGGGCTCGGCTCGGCTCGGCTCGGCTCGGCTC 4320
 4321 GACCTGCTACTCTCTGCAATGCGTGAATTCAGCTGAGGAAATGATGATGATGATGATGAT 4380
 4321 GACCTGCTACTCTCTGCAATGCGTGAATTCAGCTGAGGAAATGATGATGATGATGATGAT 4380
 4381 GCTGAGGTAGAAATCATCCCGGCGCGGCGCTTGTGTAAACCGCTTCCAGACACT 4440
 4381 GCTGAGGTAGAAATCATCCCGGCGCGGCGCTTGTGTAAACCGCTTCCAGACACT 4440
 4441 GCCAGCACCGTATGTCAGAGGTGATCTGTAGACAAACGAGTCTGCAATGCAAAAGAA 4500
 4441 GCCAGCACCGTATGTCAGAGGTGATCTGTAGACAAACGAGTCTGCAATGCAAAAGAA 4500
 4501 AACCAAGGTGCGGAGAAAGACCTCTGAGCCCTCTGTACACGAGCTCTTGTCTTCCAGGA 4560
 4501 AACCAAGGTGCGGAGAAAGACCTCTGAGCCCTCTGTACACGAGCTCTTGTCTTCCAGGA 4560
 4561 GAGCCCTCCAGGAGGAGGTGTTACAGATCATCTGCTGAGGAGATTCAGTTCAGATCA 4620
 4561 GAGCCCTCCAGGAGGAGGTGTTACAGATCATCTGCTGAGGAGATTCAGTTCAGATCA 4620
 4621 CAAATCTTTATGAGGAGTGGCCCGAGTACTCTTGTAGACAAACGAGTCTGCAATGCAAA 4680
 4621 CAAATCTTTATGAGGAGTGGCCCGAGTACTCTTGTAGACAAACGAGTCTGCAATGCAAA 4680
 4681 GATTGAGTGTTCAAACTCTTCT 4740
 4681 GATTGAGTGTTCAAACTCTTCT 4740
 4741 GACAAAGAGAGTTCCTCCAAATGCTCTCTGAGAAAGTCTTACCGGACCTTCTTACTCTGTT 4800
 4741 GACAAAGAGAGTTCCTCCAAATGCTCTCTGAGAAAGTCTTACCGGACCTTCTTACTCTGTT 4800
 4801 ATAGCAACTATTAACACTGTTGTCACAAACAGCGATACAAACAGGAGAAACGCA 4860
 4801 ATAGCAACTATTAACACTGTTGTCACAAACAGCGATACAAACAGGAGAAACGCA 4860
 4861 CAGGTGGAAGCCCTCTGTTAAACAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4920
 4861 CAGGTGGAAGCCCTCTGTTAAACAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4920
 4921 GGGATACCAAGCAGTCTGTTGTTCTTCTGAGGAGTCTGATGATGATGATGATGATGATGAT 4980
 4921 GGGATACCAAGCAGTCTGTTGTTCTTCTGAGGAGTCTGATGATGATGATGATGATGATGAT 4980

RESULT 3

AB021131

LOCUS

DEFINITION

AB021131

ACCESSION

AB021131.1

VERSION

GI:11611474

KEYWORDS

Source

ORGANISM

Mus musculus

Mus musculus

Bukaryota; Eukarya; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

1 (sites)

REFERENCE

AUTHORS

AB021131 4800 bp mRNA linear ROD 20-DEC-2000
 Mus musculus RIM2 mRNA, complete cds.

AB021131.1 GI:11611474

Rim2.

Mus musculus (house mouse)

Mus musculus

Bukaryota; Eukarya; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

1 (sites)

REFERENCE

AUTHORS

Ozaki, N., Shibasaki, T., Kashima, Y., Miki, T., Takahashi, K., Ueno, H.,

Sunaga, Y., Yano, H., Matsuura, Y., Iwanaga, T., Takai, Y. and Seino, S.

TITLE	CAMP-GEFII is a direct target of cAMP in regulated exocytosis
JOURNAL	Nat. Cell Biol. 2 (11), 805-811 (2000)
MEDLINE	20512528
PUBMED	11056535
REFERENCE	2 (bases 1 to 4800)
AUTHORS	Ozaki, N. and Seino, S.
TITLE	Direct Submission
JOURNAL	Submitted (12-DEC-1998) Nobuaki Ozaki, Nagoya University School of Medicine, First Department of Internal Medicine, Nagoya, Nagoya 466-8550, Japan (E-mail:n-ozaki@med.nagoya-u.ac.jp, Tel:81-52-744-2142, Fax:81-52-744-2157)
COMMENT	Sequence updated (22-Jun-1999) Sequence updated (02-Feb-2000).

FEATURES
source

gene

CDS

[illegible]

Query Match 92.5%; Score 4608.4; DB 10; Length 4800;
Best Local Similarity 96.4%; Pred. No. 0;
Matches 4799; Conservative 0; Mismatches 1; Indels 180;

1	GCTCCCTTAGGTTGGTTCGGCTCCGCAACATGTCGGCTCCGCTCCGGCCCCCGGGCCG	60
1	GCTCCCTTAGGTTGGTTCGGCTCCGCAACATGTCGGCTCCGCTCCGGCCCCCGGGCCG	60
61	CCGGCTTCCACCCCGCGGCTCTCAACTCTCCGACGCCAGATCCGGAACCTCAG	120
61	CCGGCTTCCACCCCGCGGCTCTCAACTCTCCGACGCCAGATCCGGAACCTCAG	120
121	CCACTTCACGGAAGGAGGAGAAATCATCTGGCTGTCTGATCTGTACAGAAAGA	180
121	CCACTTCACGGAAGGAGGAGAAATCATCTGGCTGTCTGATCTGTACAGAAAGA	180

1261 |||||TGTTCTTTGGCAAGCTGAACTAGAAAGATTCCAGGATTTCTGCTAGGATGGATAG 1320
 1321 ACCATCAAGGCAAGAGATCTGTATCTGAACCTAGAGCTGCAATGGAAAAACCAACGATCGTA 1380
 1321 ACCATCAAGGCAAGAGATCTGTATCTGAACCTAGAGCTGCAATGGAAAAACCAACGATCGTA 1380
 1381 TTCAATGGAAAGAACTCGAGAGCTCAGGGACAAAGATCTTATTCACAAAGGACCTCAAA 1440
 1381 TTCAATGGAAAGAACTCGAGAGCTCAGGGACAAAGATCTTATTCACAAAGGACCTCAAA 1440
 1441 TCATAGTCTCTCCACCCCTCGCGGAGCCCTATACCGCTGTATAGACCCAGACATGAGGCG 1500
 1441 TCATAGTCTCTCCACCCCTCGCGGAGCCCTATACCGCTGTATAGACCCAGACATGAGGCG 1500
 1501 CGCTGACTCCCTACGGAAACAGACACCACTTAGATCCCGAGCTCTGCTGTAGGAAACGAA 1560
 1501 CGCTGACTCCCTACGGAAACAGACACCACTTAGATCCCGAGCTCTGCTGTAGGAAACGAA 1560
 1561 GCGAGAAAAATGGAAACCACTTAAAGAAATGATTTCTTTGAGTTTCAGACGATCCGATGC 1620
 1561 GCGAGAAAAATGGAAACCACTTAAAGAAATGATTTCTTTGAGTTTCAGACGATCCGATGC 1620
 1621 AGTGAGCGCGCCCGCCACCAAGGCTCTATAATCCAAAGAGGAGGTAAATAAGCGCCAGGT 1680
 1621 AGTGAGCGCGCCCGCCACCAAGGCTCTATAATCCAAAGAGGAGGTAAATAAGCGCCAGGT 1680
 1681 TTCACCTGAGAGCTCGGAGGAGGAGCTGGCATCCACCTGAGTATACAGCTGTGATGA 1740
 1681 TTCACCTGAGAGCTCGGAGGAGGAGCTGGCATCCACCTGAGTATACAGCTGTGATGA 1740
 1741 TGTGAGCTGGAAAGCAGAGTGTGATGAGAAAGGGGACGTCAAAAAGGAAAGAA 1800
 1741 TGTGAGCTGGAAAGCAGAGTGTGATGAGAAAGGGGACGTCAAAAAGGAAAGAA 1800
 1801 AACTAGTGAGCAGGAGTTTGTGCGATTTCAACACAGCTCTGAGAGACAAAGAAAG 1860
 1801 AACTAGTGAGCAGGAGTTTGTGCGATTTCAACACAGCTCTGAGAGACAAAGAAAG 1860
 1861 GATGTACTATGTTGGCCACTCTTTGGAAGAGGATTTGGAAATGCTGTAGCCCTCAGATTAA 1920
 1921 GGAATCTAGTGTGGCCACTCTTTGGAAGAGGATTTGGAAATGCTGTAGCCCTCAGATTAA 1920
 1921 GGAATCTAGTGTGGCCACTCTTTGGAAGAGGATTTGGAAATGCTGTAGCCCTCAGATTAA 1920
 1981 TAAGCACTCTGTGACCTGGCAGCATTCAAAGATGGAGATCGCTTAATTTGGTCTGATTTT 2040
 1981 TAAGCACTCTGTGACCTGGCAGCATTCAAAGATGGAGATCGCTTAATTTGGTCTGATTTT 2040
 2041 ATTAAATAAGCGTTTAAAGATGGGAGTGTACTCTGAGACTCAGGAGCAATGCTGGCTT 2100
 2041 ATTAAATAAGCGTTTAAAGATGGGAGTGTACTCTGAGACTCAGGAGCAATGCTGGCTT 2100
 2101 AAAGTTGTAGGAGGAAAGATGATCTGAATCAGGTCTGATCTGAGGAGTGTGATCTGAG 2160
 2101 AAAGTTGTAGGAGGAAAGATGATCTGAATCAGGTCTGATCTGAGGAGTGTGATCTGAG 2160
 2161 AAAAAAGCAAGTTTGTAGCTGATCTGAGGACATCTTAGACAGCTGTAGGAGTGTGAG 2220
 2161 AAAAAAGCAAGTTTGTAGCTGATCTGAGGACATCTTAGACAGCTGTAGGAGTGTGAG 2220
 2221 ATGGAATGGAGGCTTATTCAGAGGAGCCATTTGAGGAAGTTTACCAATTTCTTGA 2280
 2221 ATGGAATGGAGGCTTATTCAGAGGAGCCATTTGAGGAAGTTTACCAATTTCTTGA 2280
 2281 ATCCAAACCTGAACCAAGATTGAGCTTGTGTTTCAAGGCCAATTTGGAGATTTCTCTAG 2340
 2281 ATCCAAACCTGAACCAAGATTGAGCTTGTGTTTCAAGGCCAATTTGGAGATTTCTCTAG 2340
 2341 AATACCTGATAGCAGCATGCAACCTGGAATCCAGTTCTAGCTCATTTGATCTCAAAA 2400

2341 AATACTGATAGCAAGCATGCAACCTGGAATCCAGTTCTAGCTCATTTGATCTCTCAAA 2400
 2401 AATGACCGCTCTCTTATATCCGTTACTCTACCATGAGTCTCTGCACTCTGAGGATCT 2460
 2401 AATGACCGCTCTCTTATATCCGTTACTCTACCATGAGTCTCTGCACTCTGAGGATCT 2460
 2461 CCGCGAGTTCTTATCTGGAAGCTTTCAATAAACTATATGTTTGA CAAGTTGTTGTCACA 2520
 2461 CCGCGAGTTCTTATCTGGAAGCTTTCAATAAACTATATGTTTGA CAAGTTGTTGTCACA 2520
 2521 GTTGTATGTTTCAATTTTGGGAGCAAGGATCTCTCCCTCCAGGAGATGAGGAGGCCAG 2580
 2521 GTTGTATGTTTCAATTTTGGGAGCAAGGATCTCTCCCTCCAGGAGATGAGGAGGCCAG 2580
 2581 GAATCTTATGTTAAGATTTTACTCTCTCCAGATAGAGTGTATAAAATAAGAGAGAAC 2640
 2581 GAATCTTATGTTAAGATTTTACTCTCTCCAGATAGAGTGTATAAAATAAGAGAGAAC 2640
 2641 AAAACAGTCAAGAAAACTTTGGAACCCAAATGGAACCGAGCTTTTATTTATCTCTCT 2700
 2641 AAAACAGTCAAGAAAACTTTGGAACCCAAATGGAACCGAGCTTTTATTTATCTCTCT 2700
 2701 CCACCGAGAGAAATTCCTGGAACGAACTCTGGAATTAACCTTTGGGATCAGCTAGGT 2760
 2701 CCACCGAGAGAAATTCCTGGAACGAACTCTGGAATTAACCTTTGGGATCAGCTAGGT 2760
 2761 TCGAGAGAGAGAGAGCGAATTTCTTAGGAGAGATTTTAAATGGAATGGAACAGCTTTGCT 2820
 2761 TCGAGAGAGAGAGAGCGAATTTCTTAGGAGAGATTTTAAATGGAATGGAACAGCTTTGCT 2820
 2821 AGATGATGAGCGGCACTGTTATAGCTGAGAGCCATGATGTCTCTCATTTGCCACTGCC 2880
 2821 AGATGATGAGCGGCACTGTTATAGCTGAGAGCCATGATGTCTCTCATTTGCCACTGCC 2880
 2881 TCGCCCTTCCCATATCTGCCCCGAGGAGCTCTCAATGAGAGAGCCCAACGCGAGGCT 2940
 2881 TCGCCCTTCCCATATCTGCCCCGAGGAGCTCTCAATGAGAGAGCCCAACGCGAGGCT 2940
 2941 GCAAGGTCGAAGAGATAAGTGAAGTGTCTGACTACACTGCGAGAGATGGCT 3000
 2941 GCAAGGTCGAAGAGATAAGTGAAGTGTCTGACTACACTGCGAGAGATGGCT 3000
 3001 GGGAGTGTGTCAGATTTATCGACACATGTCGCGGAGATTTTCAAGCTCACTGTTGCT 3060
 3001 GGGAGTGTGTCAGATTTATCGACACATGTCGCGGAGATTTTCAAGCTCACTGTTGCT 3060
 3061 GCCAGAACAGTGTCTCATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 3120
 3061 GCCAGAACAGTGTCTCATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 3120
 3121 TGTATAGAGAGGACAAAGTCAATGTTGCTGCTAGTGGCCCTCTCTCTCAAGGAATGTGGA 3180
 3121 TGTATAGAGAGGACAAAGTCAATGTTGCTGCTAGTGGCCCTCTCTCTCAAGGAATGTGGA 3180
 3181 ACAGGGGACCGAGGAGCACTGTCATCTGGGCAATTAACACATTAAGCTGAGTGTAG 3240
 3181 ACAGGGGACCGAGGAGCACTGTCATCTGGGCAATTAACACATTAAGCTGAGTGTAG 3240
 3241 ACACCGTGTATGATGACCACTACTCTTTCAGATAGAGAGGAGTTGTGAGAGCAGAGA 3300
 3241 ACACCGTGTATGATGACCACTACTCTTTCAGATAGAGAGGAGTTGTGAGAGCAGAGA 3300
 3301 TAGACAGCCATATACAGATCCAGATCAACAGAAACACGCGCTCTCTCTAGAGCGGACAC 3360
 3301 TAGACAGCCATATACAGATCCAGATCAACAGAAACACGCGCTCTCTCTAGAGCGGACAC 3360
 3361 CACCGCTCCAGATCTCTGAAAGTCTGATTAACAACTCATGAGGTCGATGCTTCTTCA 3420
 3361 CACCGCTCCAGATCTCTGAAAGTCTGATTAACAACTCATGAGGTCGATGCTTCTTCA 3420
 3421 AATGACTGGAAGATCTGCCCTCTCTTACCTTATGAGGCTCTACCTCTGAGGCTTACCGG 3480
 3421 AATGACTGGAAGATCTGCCCTCTCTTACCTTATGAGGCTCTACCTCTGAGGCTTACCGG 3480
 3481 AAGTCTCAACCTCTGAGGCTTACCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3540

Db 1641 CGTGGAGATTGAAAGCGAGAGCGTAGGTGAGAAAGGAGACA----- 1681
Qy 1801 AACTAGTACGACGAGGAGTTTGTGGGATTCTTAACACCGAGTCTGAGAGACAAGAAAG 1860
Db 1682 ----- 1681
Qy 1861 GATGTACTATGTGTGGCCACTCTTTTGGAAAGGAGATTGGAATGGTCTCAGGCTCAGATTAA 1920
Db 1682 -----TGAGTACAGCTGGTGGGAGATGCGTCTTG 1712
Qy 1921 GBACTCTGGGGTAGATACCTGTGTAGTACACAACTTTAAACGAGGAGCATAGCCATATGGA 1980
Db 1713 GCA-----TAGCAGTGAAGCATCCCAATGTC 1739
Qy 1981 TAAGCACCTGTGACCTGGCAGCCATCCAAAGATGGAGATGCGCTAAATGGTGTATTTT 2040
Db 1740 TTTGCACTCTGTGACCTGGCAGCCATCCAAAGATGGAGATGCGCTAAATGGTGTATTTT 1799
Qy 2041 ATTAAATAGCGTTTAAAGATGGAGTGTACCTCGAGACTCTGAGATTCAGAGCAATGCTGGGCTT 2100
Db 1800 ATTAAATAGCGTTTAAAGATGGAGTGTACCTCGAGATTCAGAGCAATGCTGGGCTT 1859
Qy 2101 AAAGTTGTAGGAGGAAGATGACTGAATCAGGTGCACTTTGTGCAATTTATTAACAAAGT 2160
Db 1860 AAAGTTGTAGGAGGAAGATGACTGAATCAGGTGCACTTTGTGCAATTTATTAACAAAGT 1919
Qy 2161 AAAAAAGGAGTTTACTGATCTGTAGGACATCTTAGACCAAGGTGATGAGTCTTGGGA 2220
Db 1920 GAAAAAGGAGTTTACTGATCTGTAGGACATCTTAGACCAAGGTGATGAGTCTTGGGA 1979
Qy 2221 ATGGAATGGAGGCTATTGCAAGGAGCCACATTTGAGGAAGTTTAAACATATTCTTGA 2280
Db 1980 ATGGAATGGAGGCTATTGCAAGGAGCCACATTTGAGGAAGTTTAAACATATTCTTGA 2039
Qy 2281 ATCCAAACCTGAACCAAGTTGAGCTTGTGTTCAGGCGCAATTTGAGATTTCTCTAG 2340
Db 2040 ATCCAAAGCTGAACCAAGTTGAGCTTGTGTTCAGGCGCAATTTGAGATTTCTCTAG 2099
Qy 2341 AATACCTGATAGCACGCGATGCAACTGGATCCAGTTCTAGCTCAATTTGAATCTCAAAA 2400
Db 2100 AATACCTGATAGCACGCGATGCAACTGGATCCAGTTCTAGCTCAATTTGAATCTCAAAA 2159
Qy 2401 AATGAGCGCTCTTCTATATCCGTATACCTCACCATGAGTCTCGGCAATGCTGAGGAGTGT 2460
Db 2160 AATGAGCGCTCTTCTATATCCGTATACCTCACCATGAGTCTCGGCAATGCTGAGGAGTGT 2219
Qy 2461 CCGCAGTTCTTATCTGGAAGCTTTCAATAAACTATGTTTGAACAGGTTGTCACCA 2520
Db 2220 CCGCAGTTCTTATCTGGAAGCTTTCAATAAACTATGTTTGAACAGGTTGTCACCA 2279
Qy 2521 GTTGTATGTTTCAATTTTGGGAGCAAGGATCTCCCTTCAGGAGATGGGAGGCCAAG 2580
Db 2280 GTTATATGTTTCAATTTTGGGAGCAAGGATCTCCCTTCAGGAGATGGGAGGCCAAG 2339
Qy 2581 GAATCCTTATGTTAAGATTTACTTCTTCAGATAGAGTGTATAAAATAAGAGAGAAC 2640
Db 2340 GAATCCTTATGTTAAGATTTACTTCTTCAGATAGAGTGTATAAAATAAGAGAGAAC 2399
Qy 2641 AAAAAAGTCAAGAAAACTTTTGGAAACCCAAATGGAAACAGACTTTTCAATTTATCTCTGT 2700
Db 2400 AAAAAAGTCAAGAAAACTTTTGGAAACCCAAATGGAAACAGACTTTTCAATTTATCTCTGT 2459
Qy 2701 CCACCGAGAGGAAATCCGTGACAGATGCTGGAAATACCTTTGGATTCAGAGTGTAGT 2760
Db 2460 TCACCGAGAGGAAATCCGTGACAGATGCTGGAAATACCTTTGGATTCAGAGTGTAGT 2519
Qy 2761 TCGAGAGAGAGAGGAAATCTTAGGAGAGATTTTAAATGAATTTGAAACAGCTTTGCT 2820
Db 2520 TCGAGAGAGAGAGGAAATCTTAGGAGAGATTTTAAATGAATTTGAAACAGCTTTGCT 2579
Qy 2821 AGATGATAGCGGCACTGTGTATAGCTGAGAGCCCATGATGCTCTCATTCGCCACTCC 2880
Db 2580 AGATGATAGCAACCACTGTGTATAGCTGAGAGCCCATGATGCTCTCATTCGCCACTCC 2639

Qy 2881 TCGCCCTTTCCCATATATCTGCCCGGAGCGAGCTCTCTATGAGAGAGAGCCCAACGCGCAGGCT 2940
Db 2640 TCACCTCTTCTCCATATATGTCACCGGAGACAGCTGCAATGGAGAGAGCCCAACGCGGTT 2699
Qy 2941 GCAAGGTTCGAGAGAGATTAAGTATAGTATGAGTGTCTGACTACGACTGCGAGATGCGCT 3000
Db 2700 GCAAGGTTCGAGAGAGATTAAGTATAGTATGAGTGTCTGACTATGACTGCGAGAGAGCGCT 2759
Qy 3001 GCGAGTAGTGTCAAGTATATCGACACAATGCGCGGAGTCTCTCAAGGTTCACGTTGCGCT 3060
Db 2760 GCGAGTAGTGTCAAGTATATCGACACAATGCGCGGAGTCTCTCAAGGTTCACGTTTATGCT 2819
Qy 3061 GCCAGAAAGTCAATCTATCAATCAATTCATTCCTCAACATCAGGCTCTCTCATCAGTAGA 3120
Db 2820 GCCAGAAAGTCAATCTATCAATCAATTCATTCCTCAACATCAGGCTCTCTCATCAGTAGA 2879
Qy 3121 TGTATAGAGAGAGAGAGTATGCTGCTAGTGTGCGCTAGTGTGCTCTCTCAAGGAAATGGA 3180
Db 2880 TGTATAGAGAGAGAGTATGCTGCTAGTGTGCGCTAGTGTGCTCTCTCAAGGAAATGGA 2939
Qy 3181 ACAGGGGACCCGAGGAGACAGCTGCTACTGCGCCATTAACAACAATTAGCGGATGATAG 3240
Db 2940 ACAGGGGCTTCGAGGAGACAGCTGCTACTGCGCCATTAACAACAATTAGCGGATGATAG 2999
Qy 3241 ACACCGTGTATGATGATGACCACTACTCTTTCAGATAGAGAC----- 3280
Db 3000 ACACCGTGTATGATGATGACCACTACTCTTTCAGAGAGAGACAGTCAATTTTCTCATCTACC 3059
Qy 3281 -----AGGATTTGTGAAGC 3294
Db 3060 TCGCTCTCGACACAGCGACAGCCAGTGTGATCAACAAGGATGGAAGGATTTGTGAAGC 3119
Qy 3295 AGCAGATAGACAGCCATATACAGATCCAGATCAACAGAAACAGCGCTCTCTTAGAGCG 3354
Db 3120 AGCAGATAGACAGCCATATACAGATCCAGATCAACAGAAACAGCGCTCTCTTAGAGCG 3179
Qy 3355 GACCAACACCCGCTCCAGATCTCTGAAAGTCTCTGATACAAACCTCATAGGCTCATAGCC 3414
Db 3180 GACCAACACCCGCTCCAGATCTCTGAAAGTCTCTGATACAAACCTCATAGGCTCATAGCC 3239
Qy 3415 TTTCAATATGATGAGAGATCTGCGCTCTCTTCACTGCTTATGAGGCTCTCACTCTCG 3474
Db 3240 TTTCAATATGATGAGAGATCTGCGCTCTCTTCACTGCTTATGAGGCTCTCACTCTCG 3299
Qy 3475 TACCGGGCTCTGTCAGACAGAGCCCATCAAGTACTCTCGGAAACAGAGCGAGGCGGCA 3534
Db 3300 CACTGGCTCTGTCAGACAGAGCCCATCAAGTACTCTCGGAAACAGAGCGAGGCGGCA 3359
Qy 3535 GCTTCAAGCTTTCACCAAGGAGAACATTTGGAGAGAGTGTCTATGCAATATAGAGAGAG 3594
Db 3360 GCTTCTCTCAGCTTCCACCAAGGAGAACATTTGGAGAGAGTGTCTATGCAATATAGAGAGAG 3419
Qy 3595 AAATCGCCAAATGAACTTTTACAAATACAAACAGGTAGCCGATCAGAGCCCGAGCTGGA 3654
Db 3420 AAATCGCCAAATGAACTTTTACAAATACAAACAGGTAGCCGATCAGAGTCCCGAGCTGGA 3479
Qy 3655 GCAAGATTAACCATTCGAAAGTATGCTCAGAGATGGGATCCACATAGAGGGGCGAGTCTGT 3714
Db 3480 ACAAGATTAACCATTCGAAAGTATGCTCAGAGATGGGATCCACATAGAGGGGCGAGTCTGT 3539
Qy 3715 TTCCAATAAATCTCTCGGACAGTGTATGATGTATCTCGGCTTTCAAGGACTAGTAG 3774
Db 3540 TTCTACTAAATCTCTCGGACAGTGTATGATGTATCTCGGCTTTCAAGGACTAGTAG 3599
Qy 3775 TGCTTCTGTTTACGAGCAGCAGCTCAATCTCGCTCAATCAGAGCGCGCGAGAGGAAA 3834
Db 3600 TGCTTCTGTTTACGAGCAGCAGCTCAATCTCGCTCAATCAGAGCGCGCGAGAGGAAA 3659
Qy 3835 CAGGAAATCAGTGTCTTTTACATCCAAATCCAAACAGACAGATGGCGGTGTGGGAA 3894
Db 3660 CAGGAAATCAGTGTCTTTTACATCCAAATCCAAACAGACAGATGGCGGTGTGGGAA 3719

4080	Db	TTACAGCTCAGAGGAAATCTGATCTTTCTGGGGTCCGCTCTGGCTCTGACAGCCAGATT	4139
4273	QY	CAGTGAATTTCTCTGGATGGCCCTGGGCCCTGCTCAGCTAGTGGGAGCGCAGACCTCTGGCTAC	4332
4140	Db	CAGGATTTCTTGATGGTCTGGGGCCCTGCTCAGCTAGTGGGAGCGCAGACCTCTGGGAC	4199
4333	QY	TCTTCGAATGGGTGACATTCAGGTGGGATGATGATATAAAGGGACAGCTGGAGTAGA	4392
4200	Db	TCTTCGAATGGGTGACATTCAGGTGGGATGATGATATAAAGGGACAGCTGGAGTAGA	4259
4393	QY	AATCATTCGGGGCGGGCCCTGTGTGTAAACCAAGGTTCCAGACACTGCGCAGACCGTA	4452
4260	Db	GATCATTCGGGGCTCGAGGCCCTGTGTGTAAACCAAGGTTCCAGACACTGCGCAGACCGTA	4319
4453	QY	TGTCAGAGTGATCTGTGTAGACAAAGGAGTCTGCATAGCCAAAAGAGAAACCAAGGTGGC	4512
4320	Db	TGTCAGAGTGATCTGTGTAGACAAAGGAGTCTGCATAGCCAAAAGAGAAACCAAGGTGGC	4379
4513	QY	GAGAAAGACCTGGAGCCCTGTACAGCAGCTCTTGTCCTTCGAGGAGAGCCCCAGGG	4572
4380	Db	GAGAAAGACCTGGAGCCCTGTATCAGCAGCTGTGTATCTTTTGAAGAGAGCCCCAGGG	4439
4573	QY	GAGGTGTTACAGATCAATCTCTGGGGAGATATGTGCTGATGATCACAAAATCCTTTAT	4632
4440	Db	GAAGTGTTACAGATCAATCTCTGGGGAGATATGGCGCATGATCACAAAATCCTTTAT	4499
4633	QY	GGAGTGCCCCAGATCTCTTAGATGAACCTGGAACATATCCAAATGCTGATGGATGGTT	4692
4500	Db	GGAGTGCCCCAGATCTCTTAGATGAACCTGGAACATATCCAAATGCTGATGGATGGTT	4559
4693	QY	CAAACTCTTCCCTCTTCCCTCTCTGATGATCCAACTTCGGGACCTCTGACAGAGAGCC	4752
4560	Db	CAAACTCTTCCCTCTTCCCTCTCTGATGATCCAACTTCGGGACCTCTGACAGAGAGCC	4619
4753	QY	TTCCCAATCGTCTCTGAAAGTTCTACCGGACCTCTTATCTCTCGTTCAATAGCAACTATA	4812
4620	Db	TTCCCAATCGTCTCTGAAAGTTCTACCGGACCTCTTATCTCTCGTTCAATAGCAACTATA	4679
4813	QY	AAACTGTGTACAAACCAACGAGATCAAAAAACCGAGAGAAAACCGACAGGTGGAGCC	4872
4680	Db	AAACTGTGTACAAACCAACGAGATCAAAAAACCGAGAGAAAACCGACAGGTGGAGCC	4738
4873	QY	CCGTGTAAACACTGCATGCTTGATGTGTGTCTACAGAGCCCAAGTCTAGGGATACCAAGC	4932
4739	Db	CCGTGTAAACACTGCATGCTTGATGTGTGTCTACAGAGCCCAAGTCTAGGGATACCAAGC	4798
4933	QY	AGTCCGTGTCTCAGAGGAAG---TCGTACACATTTGTCCTTACCAAGG	4980
4799	Db	CATCTGTGCTCTCAGAGGAAGCATCCGTGGCAGTGTGCTCTTACCAAGG	4850

RESULT	6
AF199322	
LOCUS	5640 bp mRNA linear
DEFINITION	Rattus norvegicus RIM2 (Rim2) mRNA, complete cds.
ACCESSION	AF199322
VERSION	AF199322.1 GI:8925857
KEYWORDS	
SOURCE	Rattus norvegicus (Norway rat)
ORGANISM	Rattus norvegicus Eukaryota; Metazoa Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE	1 (bases 1 to 5640) Wang Y., Sugita S. and Sudhof T.C. The RIM/MIM family of neuronal C2 domain proteins. Interactions with Rab3 and a new class of Src homology 3 domain proteins <i>J. Biol. Chem.</i> 275 (26), 20033-20044 (2000)
AUTHORS	
TITLE	
JOURNAL	
MEDLINE	20347919
PUBMED	10748113
REFERENCE	2 (bases 1 to 5640) Wang Y. and Sudhof T.C.

QY 2489 -----ATAAACTATGTTTGA CAAGTTGGTCAACAGTTGATAGTTAC 2532
Db TGTTCCTAGGGTTTCAGATATAAATAATGTTTGA CAAGTTGGTCAACAGTTGATAGTTAC 2339
QY 2533 AATTTTGGGAGCAAGGATCTCCCTCCAGGAGAGATGGGAGCCCAAGGATCTTATGT 2592
Db AATTTTGGGAGCAAGGATCTCCCTCCAGGAGAGATGGGAGCCCAAGGATCTTATGT 2399
QY 2593 TAAGATTTACTCTCTCCATGATAGAGTGAATAAAATAAGAGAAACAATAACAGTCAA 2652
Db TAAATTTACTCTCTCCAGAGAGAGTGAATAAAACAAGAGAAACAATAACAGTCAA 2459
QY 2653 GAAAACTTTGGAAACCAATATGAAACCAAGCTTTCAATTTATCTCTGTCCACCAAGAGA 2712
Db GAAAACTTTGGAAACCAATATGAAACCAAGCTTTCAATTTATCTCTGTCCACCAAGAGA 2519
QY 2713 ATTCCGTGACGATGCTGGAATATACCTTTGGGATCAAGCTAGAGTTCCGAGAGAGA 2772
Db ATTCCGGGAACGAATGCTGGAATATACCTTTGGGATCAAGCTAGAGTTCCGAGAGAGA 2579
QY 2773 GAGCGAATCTTAGAGAGAGATTTTAAATTTGAATTTGAAACAGCTTTGTAGATGATGACC 2832
Db AGTGAATCTTAGAGAGAGATTTTAAATTTGAATTTGAAACAGCTTTGTAGATGATGACC 2639
QY 2833 GCACTGGTATAAGCTGAGACCCATGATGTCTCTCATTTGCCACTCCCTGCCCCCTTCCC 2892
Db ACACTGGTACAAACTTCAGACCCCATGATGTCTCTCATTTGCCACTCCCTGCCCCCTTCCC 2699
QY 2893 ATATCTGCCCGAGGAGCTCATCGGAGAGGCCCAACCGCAGAGCTCCAAAGTTCGAA 2952
Db ATATATGCCACGAGAGAGCTGATGATGAGAGAGGCCCAACCGCAGAGTTCGAAAGTTCGAA 2759
QY 2953 GAGATAAGTGAAGTGAAGTGTCTGACTACGATCGAGAGATGGGTGGAGTAGTGTG 3012
Db GAGATAAGTGAAGTGAAGTGTCTGACTACGATCGAGAGATGGGTGGAGTAGTGTG 2819
QY 3013 AGATTATCGACAAATTTGGCGCGATTTTCAAAGCTTCAAGTGTCCGTTGTCGAGAACAGT 3072
Db AGATTATCGACAAATTTGGCGCGATTTTCAAAGCTTCAAGTGTCCGTTGTCGAGAACAGT 2879
QY 3073 CATGTATCAATCATGTCTACCATCATCGGATCTCTCTCATCGAGTAGATGTTATAGAG 3132
Db CATGTATCAATCAATCATGTCTACCATCATCGGATCTCTCTCATCGAGTAGATGTTATAGAG 2939
QY 3133 GACAGGTCAATGTCTGCTAGTGCCTCTCTCTCAAGAGATGTCGACAGGCGCACCG 3192
Db GACTAGGTCAATGTCTGCTAGTGCCTCTCTCTCAAGAGATGTCGACAGGCGCTTCTG 2999
QY 3193 AGGGAACGCTGTACTGGCCATTACAAACAATTTAGCCGATGAGATAGACCGTGTCTAT 3252
Db AGGGAACGCTGTACTGGCCATTACAAACAATTTAGCCGATGAGATAGACCGCGTGTCTAT 3059
QY 3253 GGATGACCACTACTCTTCAGATAGAC ----- 3280
Db GGATGACCACTACTCTTCAGATAGAC ----- 3119
QY 3281 -----AGGATTTGTAAGCAGCAGATAGACA 3306
Db -----AGGATTTGTAAGCAGCAGATAGACA 3179
QY 3307 GCATATATCAAGATTCAGATATCAAGAAACCAAGGCTCTCTCTAGAGCGGACCAACCCCG 3366
Db GCATATATCAAGATTCAGATATCAAGAAACCAAGGCTCTCTCTAGAGCGGACCAACCCCG 3239
QY 3367 CTCAGATCTCTGAAGCTCTGATATCAAACTCATGAGTCAATGCTTCTATTAATGAC 3426
Db CTCAGATCTCTGAAGCTCTGATATCAAACTCATGAGTCAATGCTTCTATTAATGAC 3299
QY 3427 TGAAGATCTGCCCTCTTCACTGCTTATGAGGTCTCACCTGCTAGCCGGTCTGT 3486
Db TGAAGATCTGCCCTCTTCACTGCTTATGAGGTCTCACCTGCTAGCCGGTCTGT 3359

QY 3487 CCAGCAAGCCCACTCAAGTACTCCCGGAAACAGAGAGAGGGGCGGACAGCTTCCACAGCT 3546
Db CCAGCAAGCTCCCTCAAGTACTCCCGGAAACAGAGAGAGGGGCGGACAGCTTCCACAGCT 3419
QY 3547 TCCACCAAGGGAACATTTGGAGAA ----- 3572
Db CCACCAAGGGAACACTGGAAGGAATGATCAAGAGAGCATGGACTCCACAGGAGAG 3479
QY 3573 -----CTGCTATGATATAGAGAGAGAAATCCCAAACTTAAACAAATACAA 3624
Db AAACTCAGGTGCTATGGAATATAGAGAGAGAAATCCCAAACTTAAACAAATACAA 3539
QY 3625 AAGGTAGCCGATCAGACCCCGAGCTGGAGCAAGATTACCAATTCGAAGTATGCTCAG 3684
Db AAGGTAGCCGATCAGATCCCGAGCTGGAAACAAGATTACCAATTCGAAGTATGCTCAG 3599
QY 3685 ATGGGATCCACATAGAGGGGCGAGTACTGTTTCCATAATCTTCGAGAGTGTATAG 3744
Db ATGGGATCCACATAGAGGGGCGAGTACTGTTTCCATAATCTTCGAGAGTGTATAG 3659
QY 3745 TGAATGATCTCGGTTTCAAGGACTAGTAGTCTCTCGTTTCAGCAGCACAAGCTACAT 3804
Db TGAATGATCTCGGTTTCAAGGACTAGTAGTCTCTCGTTTCAGCAGCACAAGCTACAT 3719
QY 3805 GTCCGTCCCAATCAGAGCGGCCCGAGAGAAACAGGAAATCAGTGTCTTTACATCCAAAT 3864
Db GTCCGTCCCAATCAGAAACCGGCCGAGAGAAACAGGAAATCAGTGTCTTTACATCCAAAT 3779
QY 3865 GCAAACAGACAGATGGGGTGTTCGGGAGAGAACTTGACCAAAAGCCAGGATTCAGTGG 3924
Db GCAAACAGACAGATGGGGTGTTCGGGAGAGAACTTGACCAAAAGCCAGGATTCAGTGG 3839
QY 3925 AGACATGTCTCTCACTGGAGAGAAATGACCGCAGCAGTCCGACACTGCAAGTGGGCGCCCT 3984
Db AGACATGTCTCTCACTGGAGAGAAATGACCGCAGCAGTCCGACACTGCAAGTGGGCGCCCT 3899
QY 3985 GGGTACAGTGCAAGAGCGCGGCTTAGCATTTGGGCGCAAAATGTTAGCTATTGTTGG 4044
Db GGGCACCAGTGCAAGAGCGCGGCTTAGCATTTGGGCGCAAAATGTTAGCTATTGTTGG 3959
QY 4045 TCTCTCAAGGAAAGTTCGAGTGCCTCTCAACTCAGCGCAAAACCGAAGGAGGTTAAAA 4104
Db TCTCTCAAGGAAAGTTCGAGTGCCTCTCAACTCAGCGCAAAACCGAAGGAGGTTAAAA 4019
QY 4105 GCTACCGAGCACTGTTTCAGAGAGACGCGAGACCGGGCTAGCAGTGGAGATCAGGAACTG 4164
Db GCTACCGAGCACTGTTTCAGAGAGACGCGAGACCGGGCTAGCAGTGGAGATCAGGAACTG 4079
QY 4165 GATGACCCGCGCAGCGCCGAGATTCACAGATGCGAGCATGAACAGCTATAGCTCGGA 4224
Db GATGACAGAGCAGCGCCGAGATTCACAGATGCGAGCATGAACAGCTATAGCTCGGA 4139
QY 4225 AGGAATCTGATCTTCCCTGGGGTCCGCTGCGCTTGAAGCGAGCTTCAAGTGTCTCT 4284
Db AGGAATCTGATCTTCCCTGGGGTCCGCTGCGCTTGAAGCGAGCTTCAAGTGTCTCT 4199
QY 4285 GATGCTCTGGGCGCTTCTGCTAGTGGAGCGCGAGACCTTGGCTTCTCTGCAATGGG 4344
Db TGAATGCTCTGGGCGCTTCTGCTAGTGGAGCGCGAGACCTTGGCTTCTCTGCAATGGG 4259
QY 4345 TGAATTTAGTGGGAAATGATGATATAAAGGGAAGGAGCTGGAGTGAATCATTCGGGC 4404
Db TGAATTTAGTGGGAAATGATGATATAAAGGGAAGGAGCTGGAGTGAATCATTCGGGC 4319
QY 4405 GCGCGCTTGTGTGTAACCAAGGTTCCAGAGCACTGCCAGCACTGATGTCAGAGTGA 4464
Db TCGAGGCTTGTGTGTAACCAAGGTTCCAGAGCACTGCCAGCACTGATGTCAGAGTGA 4379
QY 4465 TCTGTTAGCAACGAGTCTGATGATGCAAAAGGAAACCAAGGTTGGCGGAGAGCCCT 4524
Db CTGTTAGCAACGAGTCTGATGATGCAAAAGGAAACCAAGGTTGGCGGAGAGCCCT 4439
QY 4525 GGAGCCCTGTGTACAGCAGCTCTTGTCTTCGAGGAGAGAGCCCGGAGGAGGTTGTACA 4584

1508 TCCTACGGAACAGACCACTTAGATCCAGCTCTGCTGTGAGGAAACGAGCGGAA 1567
 831 TCCTACGGAACCAACCACTTAGATCCAGCTCTGCTGTGAGGAAACGAGCGGAA 890
 1568 AAAATGGAACCAATGTAAAGCAATGATCTTTGAGTTTCAGACCAAGTCCAGAGTCAGTGAG 1627
 891 AAAATGGAACCAATGTAAAGCAATGATCTTTGAGTTTCAGACCAAGTCCAGAGTCAGTGAG 950
 1628 CGGCCCCCAACAGGCTCATATAATCCAAAGAAAGGAGTAAATATGCCAGAGTTTCACTG 1687
 951 CGGCCCCCAACAGGCTCATATAATCCAAAGAAAGGAGTAAATATGCCAGAGTTTCACTG 1010
 1688 AGCAGCTTCGAGGAGGAGCTGCGATCCACACTGAGTATACAGCTGTGATGATGAG 1747
 1011 AGCAGCTTCGAGGAGGAGTTCGATCCACGCTGAGTATACAGCTGTGATGAGCTGAG 1070
 1748 CTGGAAGCGAGAGTGTGAGTGAAGAGGCGACAGTCAAAAGCGAAAGAAAGAAACTAGT 1807
 1071 ATTGAAGCGAGGCTGAGTGAAGAGGAC 1104
 1808 GAGCAGGAGTTTGTGCGATTTCTAAACCAAGTCTGAGAGACAAAGAAAGAGATGTAC 1867
 1105 ----- 1104
 1868 TATGTGGCCACTCTTTGGAAGAGGATTTGGAATGGTCTGAGCTCAGATTAAAGACTCT 1927
 1105 -----TGGAGTACAGCTGGTTGGAGCATGCTTGG----- 1136
 1928 GGGGTAGATACCTGTAGTAGCAACCCCTTAACGAGGAGCATAGCCATAGTATAGACAC 1987
 1137 -----CATAGCAGTGGGCATCCCAATGTCTTTGCAC 1169
 1988 CTTGTGACTGGCAGCATCCAAAGATGAGATGCGCTTAATGGTCTGATTTTATTAAT 2047
 1170 CTTGTGACTGGCAGCATCCAAAGATGAGATGCGCTTAATGGTCTGATTTTATTAAT 1229
 2048 AAGCGTTTAAAGATGGGAGTGTACTCTGAGACTCAGGAGCAATGCTGGCTTAAAGTT 2107
 1230 AAGCGTTTAAAGATGGGAGTGTACTCTGAGATTCAGGAGCAATGCTGGCTTAAAGTT 1289
 2108 GTAGGAGGAAGATGACTGAATCAGGTGCACTTTGTGCAATTTATTAACCAAGTAAAGAA 2167
 1290 GTAGGAGGAAGATGACTGAATCAGGTGCACTTTGTGCAATTTATTAACCAAGTAAAGAA 1349
 2168 GGAAGTTTATGCTGATCTGTAGGACATCTTACGAGGATGAGTCTTGGATGGAAT 2227
 1350 GGAAGTTTATGCTGATCTGTAGGACATCTTACGAGGATGAGTCTTGGATGGAAT 1409
 2228 GGGAGCTATTGCAAGGAGCCATTTGAGGAGTTTACCAATTTCTAGATTCGAA 2287
 1410 GGAAGCTATTGCAAGGAGCCATTTGAGGAGTTTACCAATTTCTAGATTCGAA 1469
 2288 CTTGAACCAAGTTGAGCTGTGTTTCAAGGCCAATTTGAGATATCTTGAATACCT 2347
 1470 CTTGAACCAAGTTGAGCTGTGTTTCAAGGCCAATTTGAGATATCTTGAATACCT 1529
 2348 GATAGCAACCATGCACTGGAATCCAGTTCTAGCTCATTTTGAATCTCAAAAATGGAAT 2407
 1530 GATAGCACTCATGCACTGGAATCCAGTTCTAGCTCATTTTGAATCTCAAAAATGGAAT 1589
 2408 GGTCTCTCTATATCCGTTACTCACCATGAGTCTGCGATGCTGAGGAGTGTCCGAG 2467
 1590 GGTCTCTCTATATCCGTTACTCACCATGAGTCTGCGATGCTGAGGAGTGTCCGAG 1649
 2468 TTCTTATCTGACAGCTTTCA----- 2488
 1650 TTCTTATCTGACAGCTTTCAAGCCAAAGCTTAGTAGAAGAAACAGCCCTTTTGTCT 1709
 2489 -----ATAAAATATGTTTGAAGGTTGTGACCAAGTTGATAGTATCAATTTTG 2539
 1710 AGGGTTCAAGTAAATATGTTTGAAGGTTGTGACCAAGTTGATAGTATCAATTTTG 1769
 2540 GGAGCAAAAGGATCTCCCTTCCAGGGAAGATGGAGGCCAAGGAATCCTTATGTTAAGATT 2599

1770 GGAGCAAAAGGATCTCCCTTTCAGGAGAGATGGAGGCCAAGGAATCTTATGTTAAATTT 1829
 2600 TACTTCTTCCAGATAGAGTGTATAAAATAGAGAGAAACAAAAAAGTCAAGAAACT 2659
 1830 TACTTCTTCCAGATAGAGTGTATAAAATAGAGAGAAACAAAAAAGTCAAGAAACT 1889
 2660 TTGGAACCCAAATAGGAACAGACTTTTCAATTTATTTCTCTGTCCACCGAAGAGATTCGT 2719
 1890 TTGGAACCCAAATAGGAACAGACTTTTCAATTTATTTCTCTGTTCACCCAGAGATTCGG 1949
 2720 GAACCAATGCTGGAATTTACCTTTGGGATCAAGCTAGAGTTCGAGAGAGAGAGAGCGAA 2779
 1950 GAACCAATGCTGGAATTTACCTTTGGGATCAAGCTAGAGTTCGAGAGAGAGAGAGTGA 2009
 2780 TTCTTAGAGAGAGATTTAAATGGAATTTGGAACAGCTTTTGTAGATGATGAGCGCACTGG 2839
 2010 TTCTTAGAGAGAGATTTAAATGGAATTTGGAACAGCTTTTGTAGATGATGAGCACTGG 2069
 2840 TATAAGCTGCGAGACCAATGATGTCTCTCATTTGCACTCTCTCTGCGCTTTCCCTATCTG 2899
 2070 TACAACCTTCAGACCAATGATGTCTCTCATTTGCACTCTCTCTCTCTCTCTCTCTCT 2129
 2900 CCCCAGGCGAGCTCCATGAGAGAGGCCCAACGCGCAGGCTCAAGGTTCGAGAGAGATA 2959
 2130 CACGAGAGAGAGCTGATGAGAGAGGCCCAACGCGCAGGCTTCAGAGTTCGAGAGAGATA 2189
 2960 AGTGAAGTGAAGTGTCTGACTCACTGCGAGGATGGCGTGGAGTGTGATGATGATGAT 3019
 2190 AGTGAAGTGAAGTGTCTGACTCACTGCGAGGATGGCGTGGAGTGTGATGATGATGAT 2249
 3020 CAGACATATGGCGCGAGTCTTCAAGAGCTCCAGTTGTGCTGCGCAGAGACAGATCATGTCA 3079
 2250 CAGACATATGGCGCGAGTCTTCAAGAGCTCCAGTTGTGCTGCGCAGAGACAGATCATGTCA 2309
 3080 TCAATATCTGCTCAGCCTCAGGCTCTCTCATGAGTGTGATGATGATGATGATGATGAT 3139
 2310 TCAATATCTGCTCAGCCTCAGGCTCTCTCATGAGTGTGATGATGATGATGATGATGAT 2369
 3140 TCAATGCTGCTGAGTGGCTCTCTCATGAGTGTGATGATGATGATGATGATGATGATGAT 3199
 2370 TCAATGCTGCTGAGTGGCTCTCTCATGAGTGTGATGATGATGATGATGATGATGATGAT 2429
 3200 CGTGCTATGCGCATATACCAACATAGCAGATGATGATGATGATGATGATGATGATGATGAT 3259
 2430 CGTGCTATGCGCATATACCAACATAGCAGATGATGATGATGATGATGATGATGATGATGAT 2489
 3260 CACTACTCTTCAGATAGAGAC----- 3280
 2490 CATTACTCTTCAGAGAGAGACAGTCAATTTTCTCACTTCACTCTGCTCTCTGACACAGGCGAG 2549
 3281 -----AGGGATTTGAGAGCAGAGATGATGATGATGATGATGATGATGATGATGATGAT 3313
 2550 ACCAGTGAACATCAACACAGGAGTGGAGGATTTGAGAGCAGAGATGATGATGATGATGAT 2609
 3314 CAGAGATCAGATATACAGAAACAAAGCTCTCTCTAGAGGAGAGACCAACCTGCTCTCAGA 3373
 2610 CAGAGATCAGATATACAGAAACAAAGCTCTCTCTAGAGGAGAGACCAACCTGCTCTCAGA 2669
 3374 TCCTCTGAGAGCTGTGATACAAACCTCATGAGTGTGATGATGATGATGATGATGATGATGAT 3433
 2670 TCCTCTGAGAGCTGTGATACAAACCTCATGAGTGTGATGATGATGATGATGATGATGATGAT 2729
 3434 TCTGCGCTCTCTCTCACTGCTTTATGAGAGTCTCACTCTCTGATACCGGCTCTCTCAGACA 3493
 2730 TCTGCGCTCTCTCTCACTGCTTTATGAGAGTCTCACTCTCTGATACCGGCTCTCTCAGACC 2789
 3494 AGCCCATCAAGTACTTCGGGAAACAGAGAGAGGCGCGAGCTTCTCAGCTTCTCCACCA 3553
 2790 AGTCCGTCAAGTACTTCGGGAAACAGAGAGAGGCGCGAGCTTCTCTCAGCTTCTCCACCA 2849
 3554 AAGGGAACATTCGAGAGAA----- 3572

QVLSSEBELASTPTTSCDDVIESSVSEKSDSOKKRNKTSBOAVNSDNTSRER
 QKVMYSGHSLDLEWSEPTQDVGDTCSSTLNEERSHSDHEPTWQPSKXGDR
 LTRILNLRKLDKSPVPSGAMGLAVGVGRTSGLCAFIITKVKGLSLADTVAHL
 RPDVRLVKNRDLQCATVEEYVNIILEKSPVVELVSVPIGMDRIPDSTHAGLE
 SSSSPESOKMDRPSISVTPMSPGMLRDVPOPLSGLSILKMFDRKGLHVLITLGA
 KOLPSRDRPNPVYKLYPLDRSDENKERTKTKLEPNWQYFVYSPVREPR
 KRMLETLMQAKRRESEFGEHILLETALDDDEPHWYKLIOTHDVSESLPSP
 YMRQLHSEPTRELRKRSISDSVSDYDCEGVDGVSDYHHRDLOSLSLYPE
 QVMSNHSPSPSGSHFLTLPRSHRQTSBHHDRDCEAADROPVHRSRSTORPL
 HRVMDHYSKSDSHFLTLPRSHRQTSBHHDRDCEAADROPVHRSRSTORPL
 RTTTSRSESRADYLMSPSLMTRSPSPALSRSPKPTGVSOTSPSTPTVGR
 RQOLPOLPPKTLERGGKLLSVTORTSTGLAVENRNTROASRSTGTSWNSY
 SEENLIFPGVRLASQSDFLDLGPAQLVGRQTLATPAMGDIVQWMDKQLEV
 EIRARGLVWPGSKTLAPYKVVILLDGVCIKKTIVARTKLEPLFLQOLLSFEES
 POKVLIIVWQDGRMDKSPMGVAQILLDELSNNVIGWFKLPPPSLSLVDPTLAP
 LTRASQSLSESTGPSYBRS*

BASE COUNT 1555 a 1190 c 1286 g 1141 t

ORIGIN

Query Match 64.2%; Score 3197.6; DB 10; Length 5172;
 Best Local Similarity 81.2%; Pred. No. 0;
 Matches 4101; Conservative 0; Mismatches 239; Indels 710; Gaps 5;

QY 1 GCTTCCCTAGGTTGGCTCGCCCAACATGTGGCTCGCTCGGGCCCCGGGCGG 60
 DB 21 GCTTCCCTAGGTTGGCTCGCCCAACATGTGGCTCGCTCGGGCCCCGGGCGG 80
 QY 61 CCGGGCTCCCAACCCCGGGGCTCTCAACCTCTCGCGAGCCCGAGATGCGGACCTCAG 120
 DB 81 CCGGGCTCCCAACCCCGGGGCTCTCAACCTCTCGCGAGCCCGAGATGCGGACCTCAG 140
 QY 121 CCACTCTACGGAAGAGAGAGAGAAATCATCTGGCTGTGATGATGCTGAGAGAGAA 180
 DB 141 CCACTCTACGGAAGAGAGAGAGAAATCATCTGGCTGTGATGATGCTGAGAGAGAA 200
 QY 181 AGAGGAG 240
 DB 201 AGAGGAG 227
 QY 241 GTGGTTTCCCTTATGTTGGGATCACTGAATGTTGAATAAAGTTCTGCGAGCCCCGAGAAA 300
 DB 228 ----- 227
 QY 301 ACAACCAATAG 360
 DB 228 -----AAGCTGCATCAACATTTGGAATGTATAAGGA 260
 QY 361 GCAAGTCAG 420
 DB 261 GCAAGTCAG 320
 QY 421 GACCTGTGGCATCTGCCAAGAGCAAAATTTGAGATGAGATGAGATGAGATGAGATG 480
 DB 321 GACCTGTGGCATCTGCCAAGAGCAAAATTTGAGATGAGATGAGATGAGATGAGATG 380
 QY 481 TTGCCAAAACCAAGTTCTGTCTGTGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 540
 DB 381 TTGCCAAAACCAAGTTCTGTCTGTGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 440
 QY 541 TATGTGGTGTGTAATTTGTCGGAAGCAAAACCAAGAAATCTCTCAATATCAGAGCATG 600
 DB 441 TATGTGGTGTGTAATTTGTCGGAAGCAAAACCAAGAAATCTCTCAATATCAGAGCATG 500
 QY 601 GTTTTAAATAGTGGTGTAAACACCTGAGAGACCTGATCAAAAGGTCTCTCAGGGCT 660
 DB 501 GTTTTAAATAGTGGTGTAAACACCTGAGAGACCTGATCAAAAGGTCTCTCAGGGCT 560
 QY 661 TCGAATATAG 720
 DB 561 TCGAATATAG 620
 QY 721 AG 780

DB 621 GGGACCCCGGGTGTACTCATGATGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAG 680
 QY 781 AAGACAGGATACTATTAATAAATGAGATGAGATGAGATGAGATGAGATGAGATGAG 840
 DB 681 AAGACAGGATCTATTAAGATGGCTGAGATGAGATGAGATGAGATGAGATGAGATGAG 740
 QY 841 TTACAGACAG 900
 DB 741 TTACAGACAG 800
 QY 901 TGAAG 960
 DB 801 TGAAG 860
 QY 961 TTGGGATATGCTGATGAG 1020
 DB 861 ATCAGATATGCTGATGAG 920
 QY 1021 TTTAAATTTACAGGATTTTAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
 DB 921 TTTAAATTTATAGGATTTTAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 980
 QY 1081 TGAATGAAGATGAG 1140
 DB 981 CBAAGAGATGAG 1040
 QY 1141 GGCACGCTACAG 1200
 DB 1041 GGCACGCTACAG 1100
 QY 1201 AGAAAGATGAG 1260
 DB 1101 AGAGCAAAATGCGAGATGCGAGATGCGAGATGCGAGATGCGAGATGCGAGATGCGAG 1160
 QY 1261 TGTTCCTTTGGCAAAACGCTGAACTAGAGATGCGAGATGCGAGATGCGAGATGCGAG 1320
 DB 1161 TGTTCCTTTGGCAAAATGCGAGATGCGAGATGCGAGATGCGAGATGCGAGATGCGAG 1220
 QY 1321 ACCATCAAGGCAAAAGATGCTGATCTGAACTAGAGATGCGAGATGCGAGATGCGAG 1380
 DB 1221 ACCATCAAGGCAAAAGATGCTGATCTGAACTAGAGATGCGAGATGCGAGATGCGAG 1280
 QY 1381 TTCAATGGAAG 1440
 DB 1281 TTCAATGGAAG 1340
 QY 1441 TCATAGTCTCTCCACCCCTCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
 DB 1341 TCATAGTCTCTCCACCCCTCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1400
 QY 1501 CGCTGATCTCCCTACGGAAG 1560
 DB 1401 TGCAGGCTCTCTACGGAAG 1460
 QY 1561 GCGAG 1620
 DB 1461 GCGAG 1520
 QY 1621 AGTGGAG 1680
 DB 1521 AGTGGAG 1580
 QY 1681 TTCACTGAG 1740
 DB 1581 TTCACTGAG 1640
 QY 1741 TGTGGAG 1800
 DB 1641 CBTGGAG 1700
 QY 1801 AACTAGTGGAG 1860
 DB 1701 AACTAGTGGAG 1760

2281 ATCCAACTGAACCAAGTTGAGTCTGTTGTTTCAAGGCCAATTGGAGATATTCCTAG 2340
2040 ATCCAAAGCCCTGAACCAAGATTGAGTCTGTTGTTTCAAGGCCAATTGGAGATATTCCTAG 2099
2341 AATACCTGATAGCAGCAGCATGCACTGGAATCCAGTTCTAGCTCAATTTGAAATCTCAAA 2400
2100 AATACCTGATAGCAGCAGCATGCACTGGAATCCAGTTCTAGCTCAATTTGAAATCTCAAA 2159
2401 AATGGACCGTCTTCTATATCCGTTACCTCACCATGAGTCTCTGGCATGCTCAGGGATGT 2460
2160 AATGGATCGTCCGCTCTATATCCGTTACCTCACCATGAGTCTCTGGCATGCTCAGGGATGT 2219
2461 CCGCAGTCTTCTATCTGGACAGCTTTCAATAAACTATGGTTTGAACAAGTTGGTCAACA 2520
2220 CCGCAGTCTTCTATCTGGACAGCTTTCAATAAACTATGGTTTGAACAAGTTGGTCAACA 2279
2521 GTTGTAGTTTACAAATTTGGGAGCAAGGATCTCCCTTCCAGGGAGATGGGAGGCCAAG 2580
2280 GTTGTAGTTTACAAATTTGGGAGCAAGGATCTCCCTTCCAGGGAGATGGGAGGCCAAG 2339
2581 GAATCCTTATGTTAAGATTTACTTCTTCCAGATAGAGTGTATATAAATAAGAGAAAGAAC 2640
2340 GAATCCTTATGTTAAGATTTACTTCTTCCAGACAGAGTGTATATAAATAAGAGAAAGAAC 2399
2641 AAAAAAGTCAAGAAAACTTTGGAAACCCAAATGGAAACCGACCTTCAATTTATCTCCTGT 2700
2400 AAAAAAGTCAAGAAAACTTTGGAAACCCAAATGGAAACCGACCTTCAATTTATCTCCTGT 2459
2701 CCACCGAAGGAATTCGGTGAACGAATGCTGGAATTTACCTTTGGGATCAAGCTAGAGT 2760
2460 TCACCGAAGGAATTCGGTGAACGAATGCTGGAATTTACCTTTGGGATCAAGCTAGAGT 2519
2761 TCAGAGAGAGAGAGAGGAGATTTCTTAGGAGAGATTTTAAATGGAATGGAAACAGCTTTGCT 2820
2520 TCAGAGAGAGAGAGAGGAGATTTCTTAGGAGAGATTTTAAATGGAATGGAAACAGCTTTGCT 2579
2821 AGATGATAGCCGCACTGGTATTAAGCTGAGACCCCAATGATGCTCTCTCATTTGCACTCCC 2880
2580 AGATGATAGAACCACTGGTATTAAGCTGAGACCCCAATGATGCTCTCTCATTTGCACTCCC 2639
2881 TCGCCCTTCCCATATCTCCCGAGGAGAGCTCCATGGAGAGAGCCCAACCGCAGGCT 2940
2640 TCACCTTCTCCCATATATGCTACGGAGACAGCTGCAATGGAGAGAGCCCAACCGCAGGTT 2699
2941 GCAGAGTCCAGAGAGATTAAGTGAAGTGTCTGACTACGACTGCGAGATGGCCT 3000
2700 GCAGAGTCCAGAGAGATTAAGTGAAGTGTCTGACTACGACTGCGAGATGGCCT 2759
3001 GGGAGTGTGTGAGATTTGCAACAATGGCCGAGTCTTCAAGCTTCAAGCTTCCGTTGGT 3060
2760 GGGAGTGTGTGAGATTTGCAACAATGGCCGAGTCTTCAAGCTTCAAGCTTCCGTTGGT 2819
3061 GCCAGAACAGTCAATGCTCAATGCTCAACATCAGGCTCTCTCATCGAGTGA 3120
2820 GCCAGAACAGTCAATGCTCAATGCTCAACATCAGGCTCTCTCATCGAGTGA 2879
3121 TGTATAGGAAGCAAGGTCATGCTGCTAGTGGCCCTCTCTCTCAAGAGGAATGGGA 3180
2880 TGTATAGGAAGCAAGGTCATGCTGCTAGTGGCCCTCTCTCTCAAGAGGAATGGGA 2939
3181 ACAGGGGCAACGAGGAGCAAGTCTGCTATGCGCCATTAACAACAATTAGCCGAATGGATAG 3240
2940 ACAGGGGCTTCGAGGGAGCAAGTCTGCTATGCGCCATTAACAACAATTAGCCGAATGGAG 2999
3241 ACACCGTGTATGATGAGCACTATCTCTTCAATAGAGAGAGAGAGGATTTGAGAGCAGAGA 3300
3000 ACACCGTGTATGAGAGCACTATCTCTTCAATAGAGAGAGAGAGGATTTGAGAGCAGAGA 3059
3301 TAGACAGCCATATACAGATCCAGATCAACAGAACACCGCCTCTCTCTAGAGCGGACAC 3360
3060 TAGACAGCCATATACAGATCCAGATCAACAGAACACCGCCTCTCTCTAGAGCGGACAC 3119

3361 CACCCGCTCCAGATCTCTGAAAGCTCTGATACAACTCTCATGAGGATGATGCTTCAAT 3420
3120 CACCCGCTCCAGATCTCTGAAAGCTCTGATACAACTCTCATGAGGATGATGCTTCAAT 3179
3421 AATGACTGAGAGATCTGCCCCCTCTCTACCTTATCGAGTCTCTACCCCTGCTAGCCG 3480
3180 AATGACTGAGAGATCTGCCCCCTCTCTACCTTATCGAGTCTCTACCCCTGCTAGCCG 3239
3481 GTCTGTCCAGACAGGCCCATCAATGATCTCCGGGAACAGAGCGAGGGCCGACAGCTTCC 3540
3240 GTCTGTCCAGACAGGCCCATCAATGATCTCCGGGAACAGAGCGAGGGCCGACAGCTTCC 3299
3541 ACAGCTTCCACCAAAAGGGAACATTTGGAGAGAGTGTCTATGAGATATAGAGAGAGAAATCG 3600
3300 TCAGCTTCCACCAAAAGGGAACATTTGGGAAG----- 3329
3601 CCAAATGAAACTTAAACAAATACAAACAGGTAGCCGATCAGAGCCCGAGCTGGAGCAAGA 3660
3330 ----- 3329
3661 TTACCAATTCGAAGTATTCGCTCAGGATGGATCCACATAGAGGGCGAGATACTGTTTCCAC 3720
3330 ----- 3329
3721 TAAATCTCTCGACAGTGTATGATGATGATCTCTGCGGTTTCAGGACTAGTAGCTTTC 3780
3330 ----- 3329
3781 TCGTTTCAGACAGCAAGCTACATGCTCCGTCCAATCAGAGCGCCCGAGAGGAAACAGGAA 3840
3330 ----- 3329
3841 AATCAGTGTCTTTTACATCCAAATGCAAAACAGAGAGATGGCGGCTGTCGGGAGAGAACTT 3900
3330 ----- 3329
3901 GACCAAAAGCACCAAGCATCAGTGGAGAGCATGTGTCTCACTGGAGAGAGATGACGGGACCA 3960
3330 ----- 3329
3961 GTCCGACACTGCGAGTGGGCGCCCTGGGTACAGTGGCAGAGAGCGGAGCTTAGCATTTGG 4020
3330 ----- 3329
4021 GCGCAAAATGTAGCTATTTGTTGTTCTCTCAGGAAAGTCCAGTCCCTCTCAACTCAG 4080
3330 ----- 3329
4081 CCAAAACCGAGAGAGAGAGGTAAAGACTACGGAGCACTGTTCAGAGAGAGCAACGGAGCCG 4140
3330 ----- AGGAGGAGGTAAAGACTACGGAGCACTGTTCAGAGAGAGCAACGGAGCCG 3380
4141 GCTAGCAGTGGAGATGAGGAACTGGAATGACCCCGCAGGCGAGCCGAGATTCAGAGTGG 4200
3381 GCTAGCAGTGGAGATGAGGAACTGGAATGACCCCGCAGGCGAGCCGAGATTCAGAGTGG 3440
4201 CAGCATGAAACAGCTATAGCTCCGAAAGGAATCTGATCTTCCTTGGGCTCCGCTTGGCTTC 4260
3441 CAGCATGAAACAGCTATAGCTCCGAAAGGAATCTGATCTTCCTTGGGCTCCGCTTGGCTTC 3500
4261 TGAACAGCAGTTCAGTGTATTCCTGATGGCTCTGGGCTCTGCTAGCTAGTGGGAGCCCA 4320
3501 TGAACAGCAGTTCAGTGTATTCCTGATGGCTCTGGGCTCTGCTAGCTAGTGGGAGCCCA 3560
4321 GACCTTGGCTACTCTGCAATGGGTGACATTCAGGTGGGATGATGATATAAAGGAGCA 4380
3561 GACCTTGGGCTACTCTGCAATGGGTGACATTCAGGTGGGATGATGATATAAAGGAGCA 3620
4381 GCTGAGGTGAGAAATCATCCGGGCGCGGCTTGTGTGTAATAAAGGTTCCAGACACT 4440
3621 GCTGAGGTGAGAGATCATCCGGGCTTGTGTGTAATAAAGGTTCCAGACACT 3680
4441 GCCAGCAACGTTATGTCAAGGTGTATCTGTTAGACCAACCGGAGTCTGCTATAGCCCAAGAA 4500

DB 441 |||||TATGTGGTGTGTATTTGTGCGAAACAAACAGAAATCCCTACCTAAGTCGGGCGCGTG 500
QY 601 GTTTTATATAGTGGTCTAAACACCTGACGACACCTGATCAAAAGTTCTCTGAGGGCT 660
DB 501 GTTTTACAAATAGTGGTCTAAATACACCGCAGCAGCTGATCAAAAGGCTCTTGGAGGGCT 560
QY 661 TCGAATAGGAGGAGCCCTCAGAGAGAGAAACCAAACTACACGACGACCCAGCTTCCA 720
DB 561 TCGCAGTAGGAGGAGCCCTCAGAGAGAGAAACCAAACTGCAATGACGACGACCGAGTTCCA 620
QY 721 AGGAGCCCGAGGTGACTTATCAGTACTCTGAGTGAAGAGCCGAGCTCAATGGGCTCAC 780
DB 621 GGGACCCCGGTGACTTATCAGTACTCTGAGTGAAGAGCCGAGCTCAATGGGCTCAC 680
QY 781 AAGACAGGATCTATTAATAATGGATCAGGAGTGAAGCAACAGATTTGCCAGTGAATGCC 840
DB 681 AAGACAGGATCTATTAATAATGGCTCAGGATGAAGCAACAGATCGCCAGTGAATGCC 740
QY 841 TTCAGACAGAAACGAGTCCATCAGTGTCCAGGAGTCAAAATCGAAGATACGACCAAG 900
DB 741 TTCAGACAGAAACGAGTCCATCAGTGTCCAGGAGTCAAAATCGAAGATACGACCAAG 800
QY 901 TGAAGAAAGAGAGGACTACTCAGTATCTCTTCCAGATGTACAAATCCCAAGATCTCC 960
DB 801 TGAAGAAAGAGAGGATATTTCCAGTATGTCTTCCAGATAGCAATCCCTAGATCTCC 860
QY 961 TTCGAGTATGCTGATAGACATCTCAGCGTGAAGCTCAATTTATGAAGAACCTGGTCA 1020
DB 861 ATCAGATATGCTGATAGACATCTCAAGTGAAGCTCAATTTATGAAGAACCTGATCA 920
QY 1021 TTTAAATTAACAGGATCTTAACAGGAGGCGATAGACATTCCAAAGATATTTGTGGA 1080
DB 921 TTTAAATTAACAGGATCTTAACAGGAGGCGATAGACATTCCAAAGATATTTGTGGA 980
QY 1081 TCAATCAAGATGTGAAGAGCAGAGATGAATATGAAGACAAAGAGAGAGAGATACCA 1140
DB 981 CGACGAGATGTGAAGAGCAGAGATGAATATGAAGACAAAGAGAGAGAGATACCA 1040
QY 1141 GGCACGCTACAGAGTGATCCAAATCTGCGCGGTATCCCGTAAAGCCCAACCCCTACGA 1200
DB 1041 GGCACGCTACAGAGTGATCCAAATTTGCGCGGTATCCCGTAAAGCCCAACCCCTACGA 1100
QY 1201 AGAAACAATGCGGATCCAGCTGAGGTGTCAGGGCAACGATGAGAGAGAGAGAGATGA 1260
DB 1101 AGAGCAAAATGCGGATCCAGCTGAGGTGTCAGGGCAACGATGAGAGAGAGAGATGA 1160
QY 1261 TGTTCCTTTGGCAACCGCTGAAGTGAAGATTCAGGATTTCTCTGCTAAGGATGATAG 1320
DB 1161 TGTTCCTTTGGCAAAATGCTGAAGTGAAGATTCAGGATTTCTCTGCTAAGGATGATAG 1220
QY 1321 ACCATCAAGGCAAGATCTGTATCTGAACGTAGAGCTGCAATGGAACCAACGATCGTA 1380
DB 1221 ACCATCAAGGCAAGATCTGTATCTGAACGTAGAGCTGCAATGGAACCAACGATCGTA 1280
QY 1381 TTCATATGAAGAACTCTGAGAGGCTCAGGCAAAAGTTCTTATCCAAGAGCTCAAA 1440
DB 1281 TTCATATGAAGAACTCTGAGAGGCTCAGGCAAAAGTTCTTATCCAAGAGGACCAAAA 1340
QY 1441 TCATAGTCTCCCAACCCCTCGGCGAGCCCTATACCGTTGATGACACGACATCAGGCG 1500
DB 1341 TCATAGTCTCTCTACCCCTCGGCGAGCCCTATACCGCTGATGACACGAGCTGAGGCG 1400
QY 1501 CGCTGACTCTCTACGGAACAGCACTTATAGTCCAGCTCTGCTGTGAGGAAACGAA 1560
DB 1401 TCGGACTCTCTACGGAACAGCACTTATAGTCCAGCTCTGCTGTGAGGAAACGAA 1460
QY 1561 GCGAGAAATAATGAACCAATGTTAGGAATGATCTTTGAGTTCAGACAGCTCCGAGTC 1620
DB 1461 GCGAGAAATAATGAACCAATGTTAGGAATGATCTTTGAGTTCAGACAGCTCTGAGTC 1520
QY 1621 AGTGAGGCGCGCCCAACGAGGCTCATTAATTCGAAGAGGAGGTAAATTCGCGCAGGT 1680
|||||

DB 1521 AGTGAGGCGCGCCCAACGAGGCTCATTAATTCGAAGAGGAGTAAATTCGCGCAGGT 1580
QY 1681 TTCATGAGGAGCTCGAGAGGAGGCTGGCAATCCACACTCAGATATACAGCTGTGATGA 1740
DB 1581 TTCATGAGGAGCTCAGAGAGGAGGTTGGCAATCCACGCTGATATACAGCTGTGATGA 1640
QY 1741 TGTGAGCTGGAAGAGGAGGAGTGTGAGTGAAGAGGAGGAGTCAAAAGGAGAAAGAAA 1800
DB 1641 CGTGAGATTCGAAGAGGAGGAGGAGTGTGAGTGAAGAGGAGCA----- 1681
QY 1801 AACTAGTGAAGAGGAGGAGTGTGAGTGTCTAAACACGAGCTCTGAGAGCAAAAGAAAAG 1860
DB 1682 ----- 1681
QY 1861 GATGTACTATGTGTGGCCACTCTTTTGGAGAGAGATTTGGATGGTCTGAGCTTCAGATTA 1920
DB 1682 -----TGGAGTACAGCTGGTTGGAGCATGGCTCTTG 1712
QY 1921 GGCATCTGGGTAGATACCTGTGTAGTGAACAACCTTAAACGAGGAGCATAGCCATAGTGA 1980
DB 1713 G-----CATAGCTAGTGAAGCATATCCCATATGTC 1739
QY 1981 TAAAGCAACCTGTGACCTGGCAGCCATCCAAAGATGGAGATCCCTTAATTTGGTGTATTTT 2040
DB 1740 TTTGCAACCTGTGACCTGGCAGCCATCCAAAGATGGAGATCCCTTAATTTGGTGTATTTT 1799
QY 2041 ATTAATAAGGCTTTAAAGATGGAGGTGTACTCTGAGACTCAGAGCAATGCTGGGCTT 2100
DB 1800 ATTAATAAGGCTTTAAAGATGGAGGTGTACTCTGAGACTCAGAGCAATGCTGGGCTT 1859
QY 2101 AAGGCTGTGAGGAGGAGGAGTGTGATGATGAGTGTGATGATGATGATGATGATGATGATG 2160
DB 1860 AAGGCTGTGAGGAGGAGGAGTGTGATGATGAGTGTGATGATGATGATGATGATGATG 1919
QY 2161 AAAAAAGCAAGTTAGCTGTATCTGTAGAGCATCTGTAGACCAAGTGTATGATGATGATG 2220
DB 1920 GAAAAAGCAAGTTAGCTGTATCTGTAGAGCATCTGTAGAGCATCTGTAGAGCATGATG 1979
QY 2221 ATGGAATGGAGGCTTATTTGCAAGGAGGCACTTTGAGGAGTTTACAACTATTTCTAGA 2280
DB 1980 ATGGAATGGAGGCTTATTTGCAAGGAGGCACTTTGAGGAGTTTACAACTATTTCTAGA 2039
QY 2281 ATCCAAACCTGAAACCAAGTTGAGCTTGTGTTTCAAGGCGCAATGAGAGATTTCTTAG 2340
DB 2040 ATCCAAAGCTGAAACCAAGTTGAGCTTGTGTTTCAAGGCGCAATGAGAGATTTCTTAG 2099
QY 2341 AATACCTGATAGCAACGATGACCACTGATGACCACTGATGACCACTGATGATGATGATG 2400
DB 2100 AATACCTGATAGCACTGATGACCACTGATGACCACTGATGATGATGATGATGATG 2159
QY 2401 AATGAGACCTCTTCTATATCTGTTTACCTCAACCAATGAGTCTCTGAGGAGTGT 2460
DB 2160 AATGAGACCTCTCTATATCTGTTTACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2219
QY 2461 CCGGAGTCTTATCTGAGCAGCTTTCAATAAATCTATGTTTGAACAAGTTTGTGTCACA 2520
DB 2220 CCCACAGTCTTATCTGAGCAGCTTTCAATAAATCTATGTTTGAACAAGTTTGTGTCACA 2279
QY 2521 GTTGTATGATTAACAATTTTGGGAGCAAGGATCTCTCTCTCAGGAGGATGAGGAGGAG 2580
DB 2280 GTTGTATGATTAACAATTTTGGGAGCAAGGATCTCTCTCTCAGGAGGATGAGGAGGAG 2339
QY 2581 GAATCTTATGTTAAGATTTTACTCTCTCTCAGAGATGAGTGTATATAAATAAGAGAGAAC 2640
DB 2340 GAATCTTATGTTAAGATTTTACTCTCTCTCAGAGATGAGTGTATATAAATAAGAGAGAAC 2399
QY 2641 AAAAAAGTCAAGAAACCTTTGGAACCCCAATGGAACCAAGCTTTTCAATTTTCTCTGT 2700
DB 2400 AAAAAAGTCAAGAAACCTTTGGAACCCCAATGGAACCAAGCTTTTCAATTTTCTCTGT 2459
QY 2701 CCACCGAGAGGATTTCCGTGAGCAAGTGTGGAATTTACCTTTGGGATTCAGCTAGGT 2760
DB 2460 TCACCGAGAGGATTTCCCGGAGCAAGTGTGGAATTTACCTTTGGGATTCAGCTAGGT 2519
|||||

QY 3181 ACAGGGGACCCGAGGGGACAGTGTCTACTGGCCATTAACAACAATAGCCGGAATGGATAG 3240
Db 2940 ACAGGGGCTTCGAGGGGACAGTGTCTACTGGCCATTAACAACAATAGCCGGAATGGATAG 2999
QY 3241 ACACGGTGTCAATGAGTACCACTTCTCTTCAATAGAGAC 3280
Db 3000 ACACGGTGTCAATGAGTACCACTTCTTCAATAGAGAGAC 3059
QY 3281 -----AGGGAATTGTGAAGC 3294
Db 3060 TCGCTCTCGACACAGGCGACAGCAGTGGAGCATCAACACAGGGATGGAGGATGTGAAGC 3119
QY 3295 AGCAGATAGACAGCCATATCAAGATCCAGATCAACAAGACCAAGCCCTCTCTCTAGAGCG 3354
Db 3120 AGCAGATAGACAGCCATATCAAGATCCAGATCAACAAGACCAAGCCCTCTCTCTAGAGCG 3179
QY 3355 GACCAACCCCGCTCCAGATCTCTGAAAGCTCTGATACAAACCTCATAGAGTCCATGCC 3414
Db 3180 GACCAACCCCGCTCCAGATCTCTGAAAGCTCTGATACAAACCTCATAGAGTCCATGCC 3239
QY 3415 TTCAATTAATGACTGGAGATCTGCCCCCTCTTCACTGCTTATCGAGGTCTCACCCCTCG 3474
Db 3240 TTCAATTAATGACTGGAGATCTGCCCCCTCTTCACTGCTTATCGAGGTCTCACCCCTCG 3299
QY 3475 TACCGGCTCTGTCAGACAGAGCCATCAAGTACTCCGGGAAACAGAGCAAGGGGCGGACA 3534
Db 3300 CACTGGCTCTGTCAGACAGAGCCATCAAGTACTCCGGGAAACAGAGCAAGGGGCGGCGCA 3359
QY 3535 GCTTCCACAGCTTCCACCAAGGGAACATTTGGAGAGAGTGTCTATGGATATAGAGGAG 3594
Db 3360 GCTTCCACAGCTTCCACCAAGGGAACATTTGGAGAGAGTGTCTATGGATATAGAGGAG 3395
QY 3595 AAATGCGCAAAATGAACCTTAACAATAACAACAGGTAGCCGGATCAGACCCCGAGACTGGA 3654
Db 3396 ----- 3395
QY 3655 GCAAGATTAACCAATTCGAAGTATGCTCAGGATGGGATCAATAGAGGGGCGAGATCTGT 3714
Db 3396 ----- 3395
QY 3715 TTCCACTAAATCTCGGACAGTGAATGAATGATGATCTGCGGTTTCAAGGACTAGTAG 3774
Db 3396 ----- 3395
QY 3775 TGCTTCTCGTTTCAGCGACCAAGCTACATGTCGTCCTCAATCAGAGCGCGCCGAGAGAA 3834
Db 3396 ----- 3395
QY 3835 CAGGAAATCAATGCTTTTATCAATCCAAATGCAGAAACAGACAGATGGCGGTGTGGGGAA 3894
Db 3396 ----- 3395
QY 3895 GAACCTTGACCAAAAGCACCAGCATCAGTGGAGACATGTCTCACTCGAGAGATGACGG 3954
Db 3396 ----- 3395
QY 3955 CAGCCAGTCCGACACTGAGTGGCGCCCTGGGTACCACTGAGTGGCAAGAGCGCGCATAG 4014
Db 3396 ----- 3395
QY 4015 CATTTGGGCGCAAAATGGTAGTATTGTTGTTCTCTACGGAAGATGCGAGTGCCTTCA 4074
Db 3396 ----- 3395
QY 4075 ACTCAGCCAAACCGAGGAGGAGTAAAGCTACGAGCAGTCTTTCAGAGAGCAGCA 4134
Db 3396 -----AGGAGGAGGTTAAAGCTACGCGACATGTCTCAGAGAGCAGCA 3440
QY 4135 GACCGGGCTAGCAGTGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4194
Db 3441 GACCGGGCTAGCAGTGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3500

QY 4195 AGATGGCGAGTGAACAGCTATAGCTCGAAGAGAAATCTGATCTTCCCTGGGGTCCGGCT 4254
Db 3501 GATGGCGAGTGAACAGCTATAGCTCGAAGAGAAATCTGATCTTCCCTGGGGTCCGGCT 3560
QY 4255 GGCCTCTGACAGCCAGTTCAGTGAATTTCTGATGGCCCTGGCCCTGCTCAGCTAGTGG 4314
Db 3561 GGCCTCTGACAGCCAGTTCAGTGAATTTCTGATGGCCCTGGCCCTGCTCAGCTAGTGG 3620
QY 4315 AGCCAGAGCCCTGGCTACTCTCTGCAATGGGTGACATTCAGGTGGGAATGATGGATAAAA 4374
Db 3621 AGCCAGAGCCCTGGCGACTCTCTGCAATGGGTGACATTCAGGTGGGAATGATGGATAAAA 3680
QY 4375 GGGCAGCTGGAGGTAGAAATCATCTCGGCGCGCGCTTGTGTGTAACACAGGTTCGA 4434
Db 3681 GGGCAGCTGGAGGTAGAAATCATCTCGGCGCGCGCTTGTGTGTAACACAGGTTCGA 3740
QY 4435 GACATGCGCAGCACCGTATGTCAGGTGTATCTGTGTAGCAACGAGTGTGTCATAGCCAA 4494
Db 3741 GACATGCGCAGCACCGTATGTCAGGTGTATCTGTGTAGCAACGAGTGTGTCATAGCCAA 3800
QY 4495 AAAGAAAACCAAGGTGGCGAGAGACCTTGGAGCCCTGTATCCAGCGCTCTTGTCTT 4554
Db 3801 AAAGAAAACCAAGGTGGCGAGAGACCTTGGAGCCCTGTATCCAGCGCTCTTGTCTT 3860
QY 4555 CGAGGAGAGCCCGCAGGCGAGGTGTACAGATCATCTGTGTGGGAGATATGGTGTAT 4614
Db 3861 CGAGGAGAGCCCGCAGGCGAGGTGTACAGATCATCTGTGTGGGAGATATGGTGTAT 3920
QY 4615 GATTCACAAATCTTTATGGAGTGGCCAGATACCTTGTGTAGTGAACCTGCACTATCCAA 4674
Db 3921 GATTCACAAATCTTTATGGAGTGGCCAGATACCTTGTGTAGTGAACCTGCACTATCCAA 3980
QY 4675 CATGTGTATGAGTGTTCAACTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4734
Db 3981 CATGTGTATGAGTGTTCAACTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4040
QY 4735 ACTCTGACAGAGAGCTTCCATCTCTCTGAAAGTTCCTAGCGACCTTCTTACTC 4794
Db 4041 ACTCTGACAGAGAGCTTCCATCTCTCTGAAAGTTCCTAGCGACCTTCTTACTC 4100
QY 4795 TGTTCATAGCACTATATAAACTGTGTGACAAACAGCGATACAAAACCAAGAA 4854
Db 4101 TGTTCATAGCACTATATAAACTGTGTGACAAACAGCGATACAAAACCAAGAA 4160
QY 4855 AACGCAAGGTGGAGCCCTGCTGTAACACTGCTGATGTTGTGTGTGTGTGTGTGTGTGTGT 4914
Db 4161 ACCACCCAGGTGG-AGACCCCTGCTGTAACACTGCTGATGTTGTGTGTGTGTGTGTGTGTGT 4219
QY 4915 GGTCTAGGATACCAAGAGTCTGTGTCTCAGAGGAG-----TCGTACACATTTGTGCC 4970
Db 4220 GGTCTAGGATACCAAGAGTCTGTGTCTCAGAGGAGCCATCCGTGTGTGTGTGTGTGTGTGTGT 4279
QY 4971 CTAGCAAAAGG 4980
Db 4280 CTAGCAAAAGG 4289

RESULT 12
AF199327

LOCUS AF199327 5031 bp mRNA linear ROD 04-JUL-2000
DEFINITION Rattus norvegicus RIM2-4A (Rim2-4A) mRNA, complete cds.
ACCESSION AF199327
VERSION AF199327.1 GI:8925867

KEYWORDS

SOURCE Rattus norvegicus (Norway rat)

ORGANISM

Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 5031)

Wang, Y., Sugita, S. and Sudhof, T.C.

The RIM/NIM family of neuronal C2 domain proteins. Interactions
with Rab3 and a new class of Src homology 3 domain proteins

Db 1281 TTCAATGGAAGAACTCGAGAGCTCAGGACAAAGTTCTTATCCAAAGGACCAAA 1340
QY 1441 TCATAGTCTCCACCCCTCGCGAGCCCTATACCGCTTGATAGACCAAGATAGAGG 1500
Db 1341 TCATAGTCTCTTACCTCGCGAGCCCTATACCGCTTGATAGACCAAGATAGAGG 1400
QY 1501 CGCTGACTCCCTACCGAAACAGCACTTATAGATCCAGCTCTGCTGTGAGAAACGAA 1560
Db 1401 TGCAGACTCCCTACCGAAACAGCACTTATAGATCCAGCTCTGCTGTGAGAAACGAA 1460
QY 1561 GCGAGAAATATGGAACCAATGTAAGATGATCTTTTGTAGTTGAGACAGTCCGAGTC 1620
Db 1461 GCGAGAAATATGGAACCAATGTAAGATGATCTTTTGTAGTTGAGACAGTCCGAGTC 1520
QY 1621 AGTGAAGCCGCCCCCAAGGCTCTATAATCCAAAGAGGAGTAAATGCGCCAGT 1680
Db 1521 AGTGAAGCCGCCCCCAAGGCTCTATAATCCAAAGAGGAGTAAATGCGCCAGT 1580
QY 1681 TTCACTGAGCAGCTCGAGAGGAGCTGGCATCCACCTGATATACAGCTGTGATGA 1740
Db 1581 TTCACTGAGCAGCTCGAGAGGAGTGGCATCCACCTGATATACAGCTGTGATGA 1640
QY 1741 TGTGAGCTGGAAGCGAGCTGTGATGAGTGAAGAGGAGCAGTCAAAAGGGAAGAA 1800
Db 1641 CGTGGAGATTGAAGCGAGGCTGTGATGAGAGGAGCA 1681
QY 1801 AACTAGTGAAGGAGGAGTTTGTGCGATTCTAACACAGGCTCTGAGAGAGCAAAAG 1860
Db 1682 ----- 1681
QY 1861 GATGACTATGTTGGGCCACTCTTTTGAAGAGGATTTGGAATGCTGTAGCCCTCAGATTAA 1920
Db 1682 -----TGGATACAGCTGGTTGGAGCATGGCTCTTG 1712
QY 1921 GCACTCTGGGTAGATACCTGTATGATAGCAACCTTTAAGAGGAGCAGTACGATAGTGA 1980
Db 1713 G-----CATGACGTGAGGAGCATCCCAATGTC 1739
QY 1981 TAAGCACCTGTGACCTGGCCAGCATCCAAAGATGAGATCGCTTAATTTGCTGATTTT 2040
Db 1740 TTTGCACTCTGTGACCTGGCCAGCATCCAAAGATGAGATCGCTTAATTTGCTGATTTT 1799
QY 2041 ATTAAATAGCGTTTAAAGATGGAGTGTACCTCCAGACTCAGGAGCAATGCTGGGCTT 2100
Db 1800 ATTAAATAGCGTTTAAAGATGGAGTGTACCTCCAGATTGAGGAGCAATGCTGGGCTT 1859
QY 2101 AAAGTTGTAGGAGGAAAGATGATCAATCAGTGAATTTGTGCAATTTATACCAAGT 2160
Db 1860 AAAGTTGTAGGAGGAAAGATGATCAATCAGTGAATTTGTGCAATTTATACCAAGT 1919
QY 2161 AAAAAAGGAGTTTATGCTGATCTGTAGGACATCTTAGACAGGTGATGAGTCTTGA 2220
Db 1920 GAAAAAGGAGTTTATGCTGATCTGTAGGACATCTTAGACAGGTGATGAGTCTTGA 1979
QY 2221 ATGGAATGGAGCTATTGCAAGAGCCCAATTTGAGGAGTTTACAAATTTCTAGA 2280
Db 1980 ATGGAATGGAGCTATTGCAAGAGCCCAATTTGAGGAGTTTACAAATTTCTAGA 2039
QY 2281 ATCCAAACCTGAACCAAGTTGAGCTTGTGTTTCAAGCCCAATTTGAGATATTTCTAG 2340
Db 2040 ATCCAAAGCTGAACCAAGTTGAGCTTGTGTTTCAAGCCCAATTTGAGATATTTCTAG 2099
QY 2341 AATACTGTATAGCAACGATGCAACCTGGAATCCAGTTCTAGCTCATTTGATCTCAAA 2400
Db 2100 AATACTGTATAGCAACGATGCAACCTGGAATCCAGTTCTAGCTCATTTGATCTCAAA 2159
QY 2401 AATGAGCTGCTTCTATATCGTTTACCTCACCCATGAGTCTCGGCACTCTCAGGAGTGT 2460
Db 2160 AATGAGTGTGCTGCTATATCGTTTACCTCACCCATGAGTCTCGGCACTCTCAGGAGTGT 2219
QY 2461 CCGCAGTTCTTATCTGGAGCAGCTTTCAATAAAACTATGTTTGAACAAGTTGGTCAACCA 2520

Db 2220 CCCACAGTCTTATCTGGACAGCTTTCAATAAACTATGTTTGAACAAGTTGGTCAACA 2279
QY 2521 GTTGTAGTTTCAATTTTGGGAGCAAGATCTCCCTTCCAGGAGATGGGAGCCAG 2580
Db 2280 GTTGTAGTTTCAATTTTGGGAGCAAGATCTCCCTTCCAGGAGATGGGAGCCAG 2339
QY 2581 GAATCCTTATGTTAAGATTTACTCTTCCAGATGAGATGATATAAATAAAGAGAGAAC 2640
Db 2340 GAATCCTTATGTTAAGATTTACTCTTCCAGACAGAGTGTATAAATAAAGAGAGAAC 2399
QY 2641 AAAAAAGTCAAGAAAACTTTGGRAACCAATGGAACCAAGCTTTCATTTATCTCTGT 2700
Db 2400 AAAAAAGTCAAGAAAACTTTGGRAACCAATGGAACCAAGCTTTCATTTATCTCTGT 2459
QY 2701 CCACGAGAGAAATTCCTGTGAACGAATCTGGAATTAACCTTTGGGATCAAGCTAGAGT 2760
Db 2460 TCACCGAGAGAAATTCCTGTGAACGAATCTGGAATTAACCTTTGGGATCAAGCTAGAGT 2519
QY 2761 TCAGAGAGAGAGAGCGGAATTTCTTAGGAGAGATTTTAAATGGAATTTGGAACAGCTTTGCT 2820
Db 2520 TCAGAGAGAGAGAGCGGAATTTCTTAGGAGAGATTTTAAATGGAATTTGGAACAGCTTTGCT 2579
QY 2821 AGATCATGAGCCGCTCTGTAATGCTGCAACCAAGCTGCTCTCTCTCTCTCTCTCTCTCT 2880
Db 2580 AGATCATGAGCCGCTCTGTAATGCTGCAACCAAGCTGCTCTCTCTCTCTCTCTCTCTCT 2639
QY 2881 TCGCCCTTCCCATATCTGCCCCGAGGAGCTCTCAATGAGAGAGCCCAACCGCAGGCT 2940
Db 2640 TCAACCTTCTCATATATGCGACGAGAGAGCTGCAATGAGAGAGCCCAACCGCAGGCT 2699
QY 2941 GCAAGGTCGAAGAGAAATAAGTGAACGTAAGTGTCTGACTAGCTCTGAGAGATGGGCT 3000
Db 2700 GCAAGGTCGAAGAGAAATAAGTGAACGTAAGTGTCTGACTAGCTCTGAGAGATGGGCT 2759
QY 3001 GGGAGTGTGTCAGATTTATGCAACAATGGCGCGCATCTTCAAGCTCAGCTTCTCTCTCT 3060
Db 2760 GGGAGTGTGTCAGATTTATGCAACAATGGCGCGCATCTTCAAGCTCAGCTTCTCTCTCT 2819
QY 3061 GCCAGAACAGTCAATGTCATCAAAATCATGCTCACCATCAGGCTCTCTCTCTCTCTCTCT 3120
Db 2820 GCCAGAACAGTCAATGTCATCAAAATCATGCTCACCATCAGGCTCTCTCTCTCTCTCTCT 2879
QY 3121 TGTATAGAGAGAGAGAGTCAATGCTGCTAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3180
Db 2880 TGTATAGAGAGAGAGTCAATGCTGCTAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2939
QY 3181 ACAGGGCACCGAGAGAGAGCTGCTCTAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3240
Db 2940 ACAGGGCACCGAGAGAGAGCTGCTCTAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2999
QY 3241 ACACCGTGTCTAGTGTACCT 3280
Db 3000 ACACCGTGTCTAGTGTACCT 3059
QY 3281 -----AGGATTTGAGAGC 3294
Db 3060 TCGCTCTCGACACAGCGAGAGCCAGTGAGCATCACCAAGGAGTGAAGGATTTGAGAGC 3119
QY 3295 AGCAGATAGAGAGCCATATCAAGATTCAGATTCAGATTCAGATTCAGATTCAGATTCAGATTC 3354
Db 3120 AGCAGATAGAGAGCCATATCAAGATTCAGATTCAGATTCAGATTCAGATTCAGATTCAGATTC 3179
QY 3355 GACCAACACCGCTCCAGATCTCTGAAAGCTCTGATACCAACCTCTGAGGAGTCTCTCTCTCT 3414
Db 3180 GACCAACACCGCTCCAGATCTCTGAAAGCTCTGATACCAACCTCTGAGGAGTCTCTCTCTCT 3239
QY 3415 TTCAATTAATGATCGGAAGATCTGCCCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3474
Db 3240 TTCAATTAATGATCGGAAGATCTGCCCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3299
QY 3475 TACCGGGTCTGTCCAGACAGCCCAATCAAGTACTCCGGGAACAGGACGAGGGGCCBACA 3534
Db 3300 CACTGGCTCTGTCCAGACAGCCCAATCAAGTACTCCGGTAACAGGACGAGGGGCCBACA 3359

SDPNLARYVTKPQYEEQWELHAEVSRARHRRHSVSLANAKLEBRSRISILLRMDRPS
RQBVSERRAMNQSASYSNERTREACQSQSYFQRTNGSPFPTFRPSGFPILDRPELHR
ADRSKQHLHDPSGAKVYTKRQKOTMLRNDLSQSDBSVRPPRPHSKKXGQGR
QVSLSSREBLASTPYRSCDDVYIEESVGERGDMYEWLEASHSSSEAFMSLHP
VTPQSKQDRLTGRILLANRKLSDVPDSDGAKGLAVGVGRTSESLCAFIYVK
KGSILADVGHLPDGVLEWNGELAQATFEVYNTILSKPFPVVELVYVSRIGDMP
RIPDSTHQALESSESQWOPDPSISVTPSPSGMLRSDVPOFLSQSLISLMDFKV
GHQIVYLLGAKDLSPREDGRPNVYKIVFLPDSDKRRRTVAKTLEPMWKLQTHD
VLSVVRHFRPEMLALELWDQARVRESESEFELILELALTALDDHEPMYKLQTHD
VSSLPFLSPSPYMPNQLGSSPSGSHRVDVIGRTSRNSPSYPPPPQNVGQLATQTRAT
DLOBSLTVSPGCVWMSNCSGSPSHRVDVIGRTSRNSQTSYDCEQGVGVSDYHDOGR
GHYNTISMDRHRVMDHVSSEKDSHFTVILFRSHRQTSSEHHRGDGCSAADQPTH
RKSSTPQLEKLTITTSRESEKADTNLMSPSLMTGRTSAPPSALSRNPNPSTVSQ
TPSSFTVTRGKQPOLPPKGTALRGQKCLRTSVQSTETGFLPAGLRNMMNRQAS
RMDZGMSNYSSEGNLLIPGVLSADBDQFSDFLDGLGPAQLVGRQTLATPAMGDLQV
GMMDKXGQLEVEITRARGLVPSGKTLPAFYVYKVVLLONGVCAIKKNTKPKRLEL
LYQLLSPFBESPOCKVQLIYWKDYGDMHKSFGVQAUILLELSELNNVYGVFKLFP
PSLSDVPTPLAPTATRRASQBSLESTGSPSYSRNS"

1503 a 1156 c 1246 g 1116 t

BASE COUNT
ORIGIN

Query Match	57.4%	Score 2860.4;	DB 10;	Length 5031;
Best Local Similarity	78.0%;	Pred. No. 0;		
Matches 3938;	Conservative	0;	Mismatches 261;	Indels 851;
				Gaps 7;

1	QY	1	GCTTCCCTAGGTGGTTCCGCTCCGCAAAATGATCGGCTCCGCTCGGGCCCCCGGGCGCG	60
21	DB	21	GCTTCCCTAGGTGGTTCCGCTCCGCAAAATGATCGGCTCCGCTCGGGCCCCCGGGCGCG	80
61	QY	61	CCGGCTCCCAACCCCGCGGCTCTCAACTCTCCGCGAGCCCGAGTCCGGGACTTCAG	120
81	DB	81	CCGGCTCCCAACCCCGCGGCTCTCAACTCTCTCCGCGAGCCCGAGTCCGGGACTTCAG	140
121	QY	121	CCACTTCAACGAGAGAGAGGAAATCATCTCGGCTGTCAATGATCGTCAAGCAAGA	180
141	DB	141	CCACTTCAACGAGAGAGAGGAAATCATCAAGGCTGTCAATGATCGTCAAGCAAGA	200
181	QY	181	AGAGGAGAGAGAGAGCTCGGTCTCAAGATCAAAAGAGAAACAACAAAGCAACAACCAACA	240
201	DB	201	AGAGGAGAGAGAGAGCTCGGTCTCAA-----	227
241	QY	241	GTGGTTTCCCTTTAGTGGGATCACTGAACTGGTAAATAAGTTCTGCGAGCCCGAGCAAA	300
228	DB	228	-----	227
301	QY	301	ACAAACCAATGAGAGAGGAGCCCGCAGACAAAGCTGCACCAACAATTTGAAATGTATAGGA	360
228	DB	228	-----AAAGCTGCATCAACAATTTGAAATGTATAGGA	260
361	QY	361	GCAAGTCAAGAACATGGGAGAGGAATCGCAGCAGCAGCAAGAGCAGAGGGGTGATGCC	420
261	DB	261	GCAAGTCAAGAAATGGGAGAGGAATCACAACAGCAGCAGAGCAGAGGGCGCGGCC	320
421	QY	421	GACCTGTGGCATCTGCCAAGACAATAATTCACAGATGAGATGGCGGCATTAATTTGTTCTTA	480
321	DB	321	GACCTGTGGCATCTGCCAAGACAATAATTCACAGATGAGATGGCGGCATTAATTTGTTCTTA	380
481	QY	481	TTGCCAAACCAAGTTCTGTCTGATGTGAGGTCCAGTCTCTTTACGCTCAACCAAGGT	540
381	DB	381	TTGCCAAACCAAGTTCTGTCTCTGTGTGAGGTTCAGTGTCAATTCGCTCAACCAAGGT	440
541	QY	541	TATGTGGTGTGTAAATTTGTGCCGAAAACAACAGAAATCCTCACTAAATCAGAGCATG	600
441	DB	441	TATGTGGTGTGTAAATTTGTGCCGAAAACAACAGAAATCCTCACTAGTTCGGGGCTG	500
601	QY	601	GTTTTAAATATAGTGGGTCTAACCACTGCAGCAACCTGATCAAAAGTTTCTTCAGGGCT	660
501	DB	501	GTTTTAAATATAGTGGGTCTAAATACACCGCAGCGCTGATCAAAAGGCTCTTCAGGGCT	560
661	QY	661	TCGAAATGAGGAAGCCCTCAGGAGAGAAAGCAAACTACACAGCAGGCCCGATTTCCA	720
561	DB	561	TCGAGTGAAGGAAGCCCTCAGGAGAGAAAGCAAACTGATCAGCAGCAGCGATTTCCA	620

Db 1682 ----- 1681
QY 1861 GATGTACTATGTGGCCACTCTTTGGAGAGAGATTGGATGTCTGAGCCTCAGATTAA 1920
Db 1682 -----TGAGTACAGCTGGTTGGAGCATGGCTCTG 1712
QY 1921 GCACTCTGGGTAGATACCTGTAGTAGCAACACCTTAAACGAGGAGCATAGCCATAGTA 1980
Db 1713 G-----CATAGCAGTGGAGCATCCCAATGTC 1739
QY 1981 TAAGCACCTGTGACCTGCGACCATCCAAAGATGAGATGCGCTAAATTTGGTGGTATTT 2040
Db 1740 TTTGACCTGTGACCTGCGACCATCCAAAGATGAGATGCGCTAAATTTGGTGGTATTT 1799
QY 2041 ATTAATAAGCGTTTAAAGATGGAGTGTACTCTGAGACTCAGAGCAATGCTGGGCTT 2100
Db 1800 ATTAATAAGCGTTTAAAGATGGAGTGTACTCTGAGACTCAGAGCAATGCTGGGCTT 1859
QY 2101 AAGGTGTAGAGGAAAGATGACTGAATCAGGTGCACTTTGTGCAATTTATACCAAGT 2160
Db 1860 AAGGTGTAGAGGAAAGATGACTGAATCAGGTGCACTTTGTGCAATTTATACCAAGT 1919
QY 2161 AAAAAAGGAGCTTTAGCTGTACTGTAGGAGCCACATTTGAGGAGGTTTACACATTTATCTAGA 2280
Db 1920 GAAAAAGGAGCTTTAGCTGTACTGTAGGAGCCACATTTGAGGAGGTTTACACATTTATCTAGA 2039
QY 2281 ATCCAAACCTGAAACCAAGATGAGCTTTGTTTCAAGGCCAATTTGAGGATTTCTCTAG 2340
Db 2040 ATCCAAAGCTGAAACCAAGATGAGCTTTGTTTCAAGGCCAATTTGAGGATTTCTCTAG 2099
QY 2341 AATACCTGATAGCAGCAGCACTGGAATCCAGTTCCTAGCTCATTGGAATCTCAAAA 2400
Db 2100 AATACCTGATAGCAGCAGCACTGGAATCCAGTTCCTAGCTCATTGGAATCTCAAAA 2159
QY 2401 AATGAGACCTCTCTATATCGTTTACTTCAACCATGAGTCTCTGCGATGCTGAGGAGTGT 2460
Db 2160 AATGAGACCTCTCTATATCGTTTACTTCAACCATGAGTCTCTGCGATGCTGAGGAGTGT 2219
QY 2461 CCGCAGTCTTATCTGGAACAGCTTTCAATAAACAATAGTTTGAACAGGTTGGTCAACA 2520
Db 2220 CCGCAGTCTTATCTGGAACAGCTTTCAATAAACAATAGTTTGAACAGGTTGGTCAACA 2279
QY 2521 GTTGTAGTTACAATTTGGGAGCAAGGATCTCCCTTCCAGGAGATGGGAGGCCAAG 2580
Db 2280 GTTGTAGTTACAATTTGGGAGCAAGGATCTCCCTTCCAGGAGATGGGAGGCCAAG 2339
QY 2581 GAATCCTTATGTTAAGATTTACTCTCTCCAGATAGAGTGTATATAAATAAAGAGAGAAC 2640
Db 2340 GAATCCTTATGTTAAGATTTACTCTCTCCAGATAGAGTGTATATAAATAAAGAGAGAAC 2399
QY 2641 AAAAAAGTCAAGAAAACTTTGGAACCCAAATGGAACAGACTTTTCAATTTATCTCTGT 2700
Db 2400 AAAAAAGTCAAGAAAACTTTGGAACCCAAATGGAACAGACTTTTCAATTTATCTCTGT 2459
QY 2701 CCACGAGAGAAATTCGTTGAGAGATGCTGGAATTTACCTTTGGATCAAGCTAGAGT 2760
Db 2460 TCACGAGAGAAATTCGTTGAGAGATGCTGGAATTTACCTTTGGATCAAGCTAGAGT 2519
QY 2761 TCAGAGAGAGAGAGAGAAATTTCTAGAGAGATTTTAAATGGAATGGAACAGCTTTGTCT 2820
Db 2520 TCAGAGAGAGAGAGAGAAATTTCTAGAGAGATTTTAAATGGAATGGAACAGCTTTGT 2579
QY 2821 AGATGATAGCCGCACTGGTATAGCTGAGACCCATGATGTCTCTCATTTGCCACTCCC 2880
Db 2580 AGATGATAGCAACACTGGTATACAACTTCAGACCCATGATGTCTCTCATTTGCCACTCCC 2639
QY 2881 TCGCCCTTCCCAATATCTGCGCGAGGAGGAGCTCTCATGAGAGAGCCCAACCGGAGGCT 2940

Db 2640 TCACCCCTTCTCCATATATATGCAACGAGACAGCTCTCATGAGAGAGCCCAACAGCAGGTT 2699
QY 2941 GCMAAGTTCGAAGAGAAATAAGTGAAGTGTCTGCAAGTGTCTGCACTAGCACTGCGAGATGGCGT 3000
Db 2700 GCMAAGTTCGAAGAGAAATAAGTGAAGTGTCTGCAAGTGTCTGCACTAGCACTGCGAGACGCGT 2759
QY 3001 GCGAGTGTCTCAGATTTACGACACAAATGCGCGCATCTTCAAGCTCTCAGCTTTGTCTGCT 3060
Db 2760 GCGAGTGTCTCAGATTTACGACACAAATGCGCGCATCTTCAAGCTCTCAGCTTTATCATG 2819
QY 3061 GCCAGAACAGTCTATGTCATCAATCATTTGCTCACCATCAGGCTCTCTCTCATCCAGTAGA 3120
Db 2820 GCCAGAACAGTCTATGTCATCAACCATTTCTCACCATCAGGCTCTCTCTCATCCAGTAGA 2879
QY 3121 TGTATATAGGAAGGACAAAGTCTATGCTGCTGCTGCTGCTCTCTCTCAAGAGAAATGGA 3180
Db 2880 TGTATATAGGAAGGACAAAGTCTATGCTGCTGCTGCTGCTCTCTCTCAAGAGAAATGGA 2939
QY 3181 ACAGGCGACCCAGGAGACAGTGTCTGCTGCTGCTGCTGCTGCTCTCTCTCAAGAGAAATGGA 3240
Db 2940 ACAGGCGCTTCCAGGAGACAGTGTCTGCTGCTGCTGCTGCTGCTCTCTCTCAAGAGAAATGGA 2999
QY 3241 ACAGGCTGTCTATGATGACCACTACTCTTTCAGATAGAGAC----- 3280
Db 3000 ACAGGCTGTCTATGACCACTACTCTTTCAGAGAGAGAGTCACTTCTCTCACTCTACC 3059
QY 3281 -----AGGGAATGTGAGC 3294
Db 3060 TCGCTCTCCACACAGGAGACAGTGTGAGCATCACACAGGAGTGAAGGATTTGAGC 3119
QY 3295 AGCAGATAGACAGCCATATACAGATCCAGATCAACAGGACAAACGCGCTCTCTAGAGCG 3354
Db 3120 AGCAGATAGACAGCCATATACAGATCCAGATCAACAGGACAAACGCGCTCTCTAGAGCG 3179
QY 3355 GACCAACACCCGCTCCAGATCTCTGAACTGCTGATACAAACCTCATGAGTGCATGCC 3414
Db 3180 GACCAACACCCGCTCCAGATCTCTGAACTGCTGATACAAACCTCATGAGTGCATGCC 3239
QY 3415 TTCAATTAATGACTGGAAGATCTGCGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3474
Db 3240 TTCAATTAATGACTGGAAGATCTGCGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3299
QY 3475 TACCGGTCTCTGAGACAGCCATCAGTACTCTGCGGAGACAGGAGGAGGAGCA 3534
Db 3300 CACTGCTCTGTCAGACAGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3359
QY 3535 GCTTCCACAGCTTCCACCAAGGAGAACTTGGAGAGAGTGTCTGATATAGAGAGAG 3594
Db 3360 GCTTCTCAGCTTCCACCAAGGAGAACTTGGAGAG----- 3395
QY 3595 AAATCGCAAAATGAACCTTAAACAAATACAAACAGGTAGCCGATCAGACCCCGACTGGA 3654
Db 3396 ----- 3395
QY 3655 GCAGATTTACCAATTCGAAGTATGCTCAGGATGGATCCACATAGAGGCGGAGACTGT 3714
Db 3396 ----- 3395
QY 3715 TTCCACTAAATCTCGGACAGTGTATGTAAGTGTATCTGCGGTTTCAAGACTAGTAG 3774
Db 3396 ----- 3395
QY 3775 TGCTTCTGTTTCAGACAGCAAGCTACATGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3834
Db 3396 ----- 3395
QY 3835 CAGGAAATCAGTGTCTTTAATTCAAATTCGAAACAGACAGATGGCGGTGTCTGCGGAA 3894
Db 3396 ----- 3395
QY 3895 GAACTTGACCAAAAGCAGCAGCATCAGTGGAGACATGTGCTCACTGAGAGAGAAATGACGG 3954
Db 3396 ----- 3395

1980 ATGGAAATGGAAGGCTATTTGCAAGAGGCCACATTTGAGGAAGTTTACAAATATTTCTAGA 2039
2281 ATCCAAACCTGAACCAACCAAGTTGAGCTTGTGTTTCAAGGCCAAATGAGAGATATTCCTAG 2340
2040 ATCCAGGCTGAACCAACCAAGTTGAGCTTGTGTTTCAAGGCCAAATGAGAGATATTCCTAG 2099
2341 AATACTGATAGACGCGATGCAACACTGAATCCAGTTCTAGCTCATTTGAAATCTCAAAA 2400
2100 AATACTGATAGACCTCATGCAACACTGAATCCAGTTCTAGCTCATTTGAAATCTCAAAA 2159
2401 AATGAGCGCTCTCTATATCGTTACCTCAACCAATGAGTCTGCGATCTGAGAGATGT 2460
2160 AATGAGTCTCGCTATATCGTTACCTCTCCATGAGTCTCGCATCTGAGGATGT 2219
2461 CCGCAGTTCTTATCTGAGCAGCTTTCA 2488
2220 CCACAGTTCTTATCTGAGCAGCTTTCAAGCCAAAGCCTTAGTAGAAGCAACGCCCTTT 2279
2489 -----ATAAACTATGTTTGAAGAGTTGTCTACAGTTGATAGTTAC 2532
2280 TGTTCCTAGGTTTCAGATAAACTATGTTTGAAGAGTTGTCTACAGTTGATAGTTAC 2339
2533 AATTTTGGGAGCAAGGATCTCCCTCCAGGGAAGATGGAGGCCAAGGAATCCTTATGT 2592
2340 AATTTTGGGAGCAAGGATCTCCCTCCAGGGAAGATGGAGGCCAAGGAATCCTTATGT 2399
2593 TAAGATTTATCTTCTTCAGATAGAGTGTATATAAATAAGAGAGAAACAAACAGTCTAA 2652
2400 TAAATTTACTTCTTCAGACAGAGTGTATATAAACAAGAGAGAAACAAACAGTCTAA 2459
2653 GAAACTTTGGAAACCAAAATGGAACAGACTTTTCAATTTATCTCTGTCACACCAAGAGA 2712
2460 GAAACTTTGGAAACCAAAATGGAACAGACTTTTCAATTTATCTCTGTCACACCAAGAGA 2519
2713 ATTCGCGAAGCAAGTCTGAAATTTACCTTTGGGATCAAGCTTGAAGTTTCGAGAGAGA 2772
2520 ATTCGCGAAGCAAGTCTGAAATTTACCTTTGGGATCAAGCTTGAAGTTTCGAGAGAGA 2579
2773 GAGCAATTTCTAGAGAGATTTTAAATGAAATGGAACAGCTTTCTAGATGATGAGCC 2832
2580 AAGTGAATCTTAGAGAGATTTTAAATGAAATGGAACAGCTTTCTAGATGATGAGCC 2639
2833 GCACTGTTAATAGTGCAGACCAATGATCTCTCATTTGCCACTCCCTGCGCCCTTCCCT 2892
2640 ACACTGTTAACAATTTCAAGCCATGATGTCTCTCATTTGCCACTCCCTCACTCTCC 2699
2893 ATATCTGCGCGAGGAGCTCCATGGAGAGAGCCCAACGCGAGGCTGCAAGGTGCA 2952
2700 ATATATGCCACCGAGACAGCTGATGAGAGAGAGCCCAACGCGAGGTTTCAAGGTGCA 2759
2953 GAGAAATAGTGAAGTGTCTGACTGACCTGCGAGGATGGCGTGGAGTGTGTGTC 3012
2760 GAGAAATAGTGAAGTGTCTGACTGACCTGCGAGGAGCGCGTGGAGTGTGTGTC 2819
3013 AGATTTACGACACAAATGGCGGAGCTTCAAGCTCCAAGTGTGCTGCTGCTGCTGCTGCT 3072
2820 AGATTTACGACACAAATGGCGGAGCTTCAAGCTCCAAGTGTGCTGCTGCTGCTGCTGCT 2879
3073 CATGTCATCAATCATTTGCTCAACATCAGGCTCTCTCATTCAGTATGATTTATAGGAG 3132
2880 CATGTCATCAATCAATTTGCTCAACATCAGGCTCTCTCATTCAGTATGATTTATAGGAG 2939
3133 GACAGGCTCATGTTCCCTTAGTCCCTCTCTCTCAAGGAATGTGGAACAGGGGACCG 3192
2940 GACTAGTCAATGTTCCCTTAGTCCCTCTCTCTCAAGGAATGTGGAACAGGGGCTTCG 2999
3193 AGGGAACATGCTGCTATGCTGCTATTAACCAATTTAGCCGAATGAGACACCGTGTCTAT 3252
3000 AGGGAACATGCTGCTATGCTGCTATTAACCAATTTAGCCGAATGAGACACCGTGTCTAT 3059
3253 GGNATGACCACTACTCTTTCAGATAGAC----- 3280

3060 GGAAGCAATTAATCTTCTCAGAGAGAGACAGATCAATTTCTCACTCTACTGCTCTCGACA 3119
3281 -----AGGAAATTTGAAGCAAGCAGATAGACA 3306
3120 CAGGCAGACCACTGAGCATCAACACAGGAGTGAAGGATTTGAGGCAAGCAGATAGACA 3179
3307 GGCATATCAGAGATCCAGATCAACAGAAACACAGGCTCTCTCTAGAGCGGACCAACCCG 3366
3180 GGCATATCAGAGATCCAGATCAACAGAAACACAGGCTCTCTCTAGAGCGGACCAACCCG 3239
3367 CTCAGAGTCTCTGAACTCTGATACAAACCTCATGAGTGTGATCCCTCTCAATTAATGAC 3426
3240 CTCAGAGTCTCTGAACTCTGATACAAACCTCATGAGTGTGATCCCTCTCAATTAATGAC 3299
3427 TGAAGATCTGCGCTCTCTCACTGCTTATCGAGGTCTCAACCTCTGTAACCGGCTCTGT 3486
3300 TGAAGATCTGCGCTCTCTCACTGCTTATCGAGGTCTCAACCTCTGTAACCGGCTCTGT 3359
3487 CCGAGCAAGCCCATCAAGTATCTCCGGGAACAGAGCAAGGGCCGACAGCTTTCACAGCT 3546
3360 CCGAGCAAGCTCCGTCAGTACTCTCCGTAAACAGAGCAAGGGCCGCGAGCTTTCACAGCT 3419
3547 TCACCAAAAGGGAACATTTGAGAGAGTGTCTATGATATAGAGGAGAGAAATGCGCAAT 3606
3420 CCACCAAAAGGGAACATTTGAGAG----- 3443
3607 GAACTTAAACAAATACAAACAGAGTAGCCGATCAGACCCAGACTGAGAGCAAGATTACCA 3666
3444 ----- 3443
3667 TTCAGATGATCGCTCAGGATGGATCCATAGAGGGGCGAGTACTGTTTCCATTAATC 3726
3444 ----- 3443
3727 CTCGACAGTGTATGATGATGATCTCGGTTCAGAGCTAGTAGTCTCTCTCTCTCT 3786
3444 ----- 3443
3787 CAGCAGCAAGCTACTATGCTGCTCAATCAGAGCGCGCGAGGGAACAGGAAATCAG 3846
3444 ----- 3443
3847 TGTCTTTTACATCCAAATATGCAAAAACAGACAGATGGCGGTGTCTCGGGAAGAACTTGAACAA 3906
3444 ----- 3443
3907 AAGCACCAGCATCAGTGGAGCATGTGCTCTACTGAGAGAGATGACGCGCAGCTGTCGA 3966
3444 ----- 3443
3967 CACTGAGTGGCGGCCCTGGGTACCAATGAGCAAGAGCGCGATCTAGCATTTGGGGCCAA 4026
3444 ----- 3443
4027 AATGCTAGTATTGTTGTTCTCTCAGGAAAGTCTCAGTGTCTCTCACTCAGCCAAAC 4086
3444 ----- 3443
4087 CGAAGGAGGATTAAGCTAAGCTACGAGCACTGTTTCAGAGAGACCGCAAGCCGGCTAGC 4146
3444 -----AGAGAGGATTAAGCTAAGCTACGAGCACTGTTTCAGAGAGACCGCAAGCCGGCTAGC 3500
4147 AGTGGAGATGAGGAACTCGATGACCGCGCAGCGCAGCGGGAATCCACAGATGCGAGAT 4206
3501 AGTGGAGATGAGGAACTCGATGACCGCGCAGCGCAGCGGGAATCCACAGATGCGAGAT 3560
4207 GAAAGCTATAGCTCGGAGGAAATCTGATCTTCTCTGGGTCTCCCTTGGCTCTTGAAG 4266
3561 GAAAGCTATAGCTCGGAGGAAATCTGATCTTCTCTGGGTCTCCCTTGGCTCTTGAAG 3620
4267 CCAGTTCAGTATTTCTGATGCGCTGCGCCCTCTCAGCTAGTGTGGAACCGACCCCT 4326
3621 CCAGTTCAGGATTTCTTGTATGTTCTGCGCCCTCTCTCAGCTAGTGTGGAACCGACCCCT 3680

481	QY	TTGCCAAAACAAGTCTTGCTCTCGATGTGTGAGGTGCGAGTGTCTTTATCGCTCAAAACAGGT	540
381	DB	TTGCCAAAACAAATTTCTGTGCTCTGTTGTGAGGTGCGAGTGTCTATTAGCTTCAAAACAGGT	440
541	QY	TATGTGGGTGTGTAAATTTGTGTGCCGAAAAAAGAAATTCCTCACTAAATTCAGGAGCATG	600
441	DB	TATGTGGGTGTGTAAATTTGTGTGCCGAAAAAAGAAATTCCTCACTAANGTGGGGCGGTG	500
601	QY	GTTTTATAATAGTGGGTCTAAACACATGTCAGCAACCTGATCAAAAGGTCTCTCGAGGGCT	660
501	DB	GTTTTACATATGTGGGTCTAATATACACCGAGAGAGCCCTGATCAAAAGGTCTTTCGAGGGCT	560
661	QY	TCGAAATGAGGAAGCCCTTCAGGAGAAAGAAAGCAAACTHACAGAGAGCGCCCAAGTTCCA	720
561	DB	TCGCACTGAGGAAGCCCTCAGGAGGAAGAGGCAAAATGCGATAGACAGAGCGAGTTCCA	620
721	QY	AGGAGCCCCAGGTGACTTATCAGTACTGTCAGATTGAGAAAGCCGAGCTCATGCGCTCAC	780
621	DB	GGGACCCCCCGGTGACTCATCAGTACTGTCGAGTTGAGAGAGGTGCGAGCTCATGCGCTCAC	680
781	QY	AAGACAGGATACTATTATTAATAATCGATCAGGAGTGAAGCACACGATGTCAGTGCACATGCC	840
681	DB	AAGACAGGATTTCTATTAGAAATGGCTCAGGAATGGAAGCACAGATGCCAGTGCACATGCC	740
841	QY	TTCAGACAGAAAAACGAAGTCCATCAGTGTTCAGGGATCAAAATCBAAGATACGAGCAAG	900
741	DB	TTCAGACAGAAAAAGAAAGTCCGTGAGTCCAGGGATCAAAATCGAAGATACGACCAAG	800
901	QY	TGAGAAAGAGAGGACTACTCAGCATATGTTCTTCAGATGGTACAAATGCCAAGATCTCC	960
801	DB	TGAAGAAAGAGAGGGAATATTCTACATATGTTCTTCAGATAGCAATGCTCTAGATCTCC	860
961	QY	TTCCGATTAATGCTGATAGACGACTCTCAGCGTAGCCCTCAATTTTATGAAAGACCTTGCTCA	1020
861	DB	ATCAGATTAATGCTGATAGACGACTCTCAACTGAGCTCAATTTTATGAAAGACCTGATCA	920
1021	QY	TTTTAAATTAACGGGATTTCTAAACAGGAGAGCCATAGACATTCCAAAGATATATTGTGA	1080
921	DB	TTTTAAATTAAGGGAATTTCTAAACAGGAGAGCCATAGACATTTCCAAAGATATATTGTGA	980
1081	QY	TGATGAGATATGTGGAGAGCGAGATGAAATATGAAGACAAAGGAGAGAGGAGGATACCA	1140
981	DB	CGACGAGATATGTGGAGAGCGAGATGAATGAATGAAGAACAAGGAGAGAGGAGGATACCA	1040
1141	QY	GGCACGCTACAGAAAGTATCAAAATCTGGCCCGGTATCCGTAAGACCAACCCCTACCA	1200
1041	DB	GGCACGCTACAGAAAGTATCAAAATTTGGCCCGGTATCCGTAAGACCAACCCCTATGA	1100
1201	QY	AGAACAAATTCGGCATCCACGCTCAGGTGTTCAGGGCACGACATGAGAGAGGCAAGTGA	1260
1101	DB	AGAGCAAAATTCGGCATCCACGCTGAAGTGTCCCGGGCACGACACAGAGAGAGGCAAGTGA	1160
1261	QY	TGTTTTCTTTGGCAAAACGCTGAACTAGAGATTTCCAGGATTTCTCTGTAAGGATGGATAG	1320
1161	DB	TGTTTTCTTTGGCAAAATGCTGAACTAGAGATTTCCAGGATTTCTCTGTAAGGATGGATAG	1220
1321	QY	ACCATCAGGCAAGATCTGTATCTGAAAGTACAGCTGCAATGCAAAAACCAACGATCGTA	1380
1221	DB	ACCATCAGGCAAGATCTGTATCTGAAAGTACAGCTGCAATGCAAAAACCAAGGATCGTA	1280
1381	QY	TTCAATGGAAGAACTCGAGAGGCTCAGGCAAAAGTTCTTATTCACAAAGGACCTCAAA	1440
1281	DB	TTCAATGGAAGAACTCGAGAGGCTCAGGCAAAAGTTCTTATTCACAAAGGACCAAA	1340
1441	QY	TCATAGTCTTCCCAACCCCTCGGGGAGGCCCTATACCGCTTGATAGACCGAGATAGGGCG	1500
1341	DB	TCATAGTCTTCTTCCCAACCCCTCGGAGAGGCCCTATACCGCTCGATAGACCGAGCTAGGGCG	1400
1501	QY	CGCTGACTCCCTACGGAACACAGACCACTTAGATCCGAGCTCTGCTGTGAGGAACGAA	1560
1401	DB	TGCCGACTCTCTACGGAACCAACACACTTAGATCCGAGCTCTGCTGTAGGAACGAA	1460
1561	QY	GCAGAAAAAATGGAACCACTGTTAAGGAATGATTTTGTAGGTTTCAGACAGTCCGAGTC	1620

[illegible]

Db 2400 AAAAAAGCTCAAGAAAACTTTGGAAACCCAAATGGAACCAAGACTTTCAATTTATTTCTCTGT 2459
Qy 2701 CCACCGAAGAGAAATTCGGTGAACGAATGCTGGAATTTACCTTTGGATCAAGCTAGAGT 2760
Db 2460 TACCGAAGAGAAATTCGGGAACGAATGCTGGAATTCACCTTTGGATCAAGCTAGAGT 2519
Qy 2761 TCGAGAAGAGAGAGCAATTCCTTAGGAGAGATTTTAAATGAAATGGAACAGCTTTGCT 2820
Db 2520 TCGAGAAGAGAGAAAGTGAATTCCTAGGAGAGATTTTAAATGAAATGGAACAGCTCTGTT 2579
Qy 2821 AGATGATAGCCGCACTGCTGATATAGCTGAGACCCATGATGCTCTCATTTCCACTCC 2880
Db 2580 AGATGATGAACCACTGCTGATCAAACTTCAGACCCATGATGCTCTCATTTCCACTCC 2639
Qy 2881 TCGCCCTTCCCATATCTCCCGGAGGAGCTCCATGAGAGAGAGCCAAACGCGAGGCT 2940
Db 2640 TCAACCTTCTCATATATGCCAGGAGAGAGCTGCAATGAGAGAGAGCCAAACGCGAGGTT 2699
Qy 2941 GCAAGGTCGAAGAGAAATAGTGAAGTGAAGTGTCTGACTACGACTCGAGAGATGGCT 3000
Db 2700 GCAAGGTCGAAGAGAAATAGTGAAGTGAAGTGTCTGACTACGAGAGAGAGGCT 2759
Qy 3001 GGGATGATGTCGATTTACGACAAATGCGCGCATCTTCAAGCTCAAGCTTCTGCT 3060
Db 2760 GGGATGCTGTGATGATTTACGACAAATGCGCGCATCTTCAAGCTCAAGCTTCTGCT 2819
Qy 3061 GCGAGAAAGCTCATGCTCATCAAAATCAATGCTCACCATCAGGCTCTCTCATCGAGTAGA 3120
Db 2820 GCGAGAAAGCTCATGCTCATCAAAATCAATGCTCACCATCAGGCTCTCTCATCGAGTAGA 2879
Qy 3121 TGTATAGGAAGAGCAAGTGTCTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3180
Db 2880 TGTATAGGAAGAGCAAGTGTCTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2939
Qy 3181 ACAGGGACCGAGGAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3240
Db 2940 ACAGGGGCTTCGAGGGAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2999
Qy 3241 ACAGGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3300
Db 3000 ACAGGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3059
Qy 3301 TAGACAGCCATATACAGATTCAGATTCAGATTCAGATTCAGATTCAGATTCAGATTCAGAT 3360
Db 3060 TCGCTCTCGACACAGGAG 3102
Qy 3361 CACCCGCTCCAGATCTCTGAAAGTCTCTGATACAAACCTCATGAGTGTGCTTCAAT 3420
Db 3103 ----- 3102
Qy 3421 AATGACTGGAAGATCTGCCCTCTCTTCACTGCTTATCGAGTCTTCACTCTGCTGCTGCTGCT 3480
Db 3103 -----GGAGGCTCTCACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3125
Qy 3481 GTCTGTCCAGACAGCCCATCAAGTACTCTCGGGAACAGAGAGAGAGAGAGAGAGAGAGAGAG 3540
Db 3126 CTCTGTCCAGACAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3185
Qy 3541 ACAGCTTCCACCAAGGGAG 3600
Db 3186 TCAGCTTCCACCAAGGGAG 3215
Qy 3601 CCAATGAAACTTAAACAAATACAAACAGGATAGCGGATCAGACCCAGAGCTGAGGACAGA 3660
Db 3216 ----- 3215
Qy 3661 TTACCATTCGAAGTATGCTCTCAGGATGCGGATCCATAGAGGGGAGAGATGCTTTCCAC 3720
Db 3216 ----- 3215
Qy 3721 TAAATCTCGACAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3780
Db 3216 ----- 3215

Qy 3781 TCGTTTCAGCAGCAACAGCTACATGTCCTGTCATCAATCAGAGCGCGCGAGAGAGAAACAGGAA 3840
Db 3216 ----- 3215
Qy 3841 AATCAGTGTCTTTTACATCCAAAATGCAAAAACAGACAGATGCGGCTGTGTCTGGGAGAGAACTT 3900
Db 3216 ----- 3215
Qy 3901 GACCAAAAGCAGCAGCATCAGTGGAGAGACATGTGCTCACTCGGAGAGAAATGACGCGCAGCCA 3960
Db 3216 ----- 3215
Qy 3961 GTCCGACACTGAGTGGCGGCCCTGCGGTACCAAGTGGCAAGAGCGGCGATCTAGCANTGG 4020
Db 3216 ----- 3215
Qy 4021 GGCCAAAATGCTAGCTATTTGTTGCTCTCTCAAGGAAAGTGGCAGTGTCTCTCAACTCAG 4080
Db 3216 ----- 3215
Qy 4081 CCAAACCGAAGAGAGAGTAAAGCTACGAGCACTGTTCAGAGAGAGAGAGAGAGAGAGAGAGAG 4140
Db 3216 -----AGAGAGAGTAAAGCTACGAG 3266
Qy 4141 GCTAGCAGTGGAGATGAGAGAACTGGAATGACCGCCAGCGCCAGCGCGGAGATCCACAGATGG 4200
Db 3267 GCTAGCAGTGGAGATGAGAGAACTGGAATGACCAAGACAGCGCCAGCGCGGAGATCCACAGATGG 3326
Qy 4201 CAGGATGAACTATAGCTCGAGAGAGAAATCTGATCTTCTCTGCGGTGCTGCTGCTGCTGCTGCT 4260
Db 3327 CAGCATGAACTATAGCTCGAGAGAGAAATCTGATCTTCTCTGCGGTGCTGCTGCTGCTGCTGCT 3386
Qy 4261 TGAAGCCAGTTCAGTGAATTTCTGATGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4320
Db 3387 TGAAGCCAGTTCAGGAGATTTCTGATGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3446
Qy 4321 GACCTTGGCTATCTCTGCAATGGGTGACATTCAGGTGGGAAATGATGATGATGATGATGATGATGAT 4380
Db 3447 GACCTTGGGAGCTCTCTGCAATGGGTGACATTCAGGTGGGAGATGATGATGATGATGATGATGAT 3506
Qy 4381 GCTGAGGTAGAAATCATCTCGGCGCGCGCTTGTGTGATGATGATGATGATGATGATGATGATGAT 4440
Db 3507 GCTGAGGTAGAGATCATCTCGGCGCTGCGAGGCTTGTGTGATGATGATGATGATGATGATGATGAT 3566
Qy 4441 GCGACACCGTATGTCAGAGTGTATCTGTAGACAAACAGAGTCTGATGATGATGATGATGATGATGAT 4500
Db 3567 GCGACACCGATATGTCAGAGTGTATCTGTAGACAAACAGAGTCTGATGATGATGATGATGATGATGAT 3626
Qy 4501 AACCAAGGTGGCGAG 4560
Db 3627 AACCAAGGTGGCGAG 3686
Qy 4561 GAGCCCGCAGGAG 4620
Db 3687 GAGCCCGCAGGAG 3746
Qy 4621 CAAATCTTTATGAG 4680
Db 3747 CAAATCTTTATGAG 3806
Qy 4681 GATGAGAT 4740
Db 3807 AATGAGAT 3866
Qy 4741 GACAG 4800
Db 3867 GACAG 3926
Qy 4801 ATAGCAACTATATAAACTGTGTGACAAACAGAGAGATGATGATGATGATGATGATGATGATGATGAT 4860
Db 3927 ATAGCAACTATATAAACTGTGTGACAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3986

Search completed: December 11, 2003, 20:44:13
Job time : 17577 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 11, 2003, 14:37:34 ; Search time 1170 Seconds
(without alignments)
11489.916 Million cell updates/sec

Title: US-09-617-099B-2

Perfect score: 4980

Sequence: 1 gcttcctaggtggttcgg.....acattgtgcctagcaagg 4980

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq 19Jun03.*

```

1: /SID81/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SID81/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SID81/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SID81/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SID81/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SID81/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SID81/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SID81/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SID81/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SID81/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SID81/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SID81/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SID81/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SID81/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SID81/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SID81/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SID81/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SID81/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SID81/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SID81/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SID81/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SID81/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SID81/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SID81/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*
25: /SID81/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4980	100.0	4980	22	Mouse Rim2-encoding
2	844.8	17.0	1302	19	Human secreted pro
3	715	14.4	6637	22	Human neuroblastom
4	398.2	8.0	7325	24	Human DNA sequence
5	377	7.6	2276	21	Human secreted pro
6	318.8	6.4	526	22	Human nervous syst
7	207.6	4.2	355	20	EST clone DM118.
8	207.6	4.2	355	20	EST clone CO618.

9	195.6	3.9	378	21	AACT74934	Human ORFX ORF489
10	188.8	3.8	7395	23	ABLI2715	Drosophila melanog
11	113.8	2.3	365	22	ABAA45437	Human breast cell
12	113.8	2.3	365	22	ABAA5935	Human foetal liver
13	113.8	2.3	365	22	ABA25602	Probe #4068 for ge
14	113.8	2.3	365	22	AAK04148	Human brain expres
15	113.8	2.3	365	22	AAK29630	Human bone marrow
16	113.8	2.3	365	22	AAI14209	Probe #4142 for ge
17	113.8	2.3	365	22	AAI35590	Probe #4276 used t
18	113.8	2.3	365	22	AAI04045	Probe #4036 used t
19	113.8	2.3	365	23	ABS29267	Human liver single
20	113.8	2.3	365	24	ABS04186	Human genome-deriv
21	111.4	2.2	366	22	AAH98466	Rat EST-derived co
22	80.4	1.6	219	22	ABA51486	Human breast cell
23	80.4	1.6	219	22	ABA36435	Probe #14901 for g
24	80.4	1.6	219	22	AAK17764	Human brain expres
25	80.4	1.6	219	22	AAK43584	Human bone marrow
26	80.4	1.6	219	22	AAI49642	Probe #18328 used
27	80.4	1.6	219	22	AAI09911	Probe #9902 used t
28	80.4	1.6	219	23	ABS43229	Human liver single
29	80.4	1.6	219	24	ABS17744	Human genome-deriv
30	76.8	1.5	169	24	ABS69183	Novel murine polyn
31	72.2	1.4	20604	23	ABLI2714	Drosophila melanog
32	69.8	1.4	153	22	ABA50565	Human foetal liver
33	69.8	1.4	153	22	ABA68523	Human breast cell
34	69.8	1.4	153	22	ABA35504	Probe #13970 for g
35	69.8	1.4	153	22	AAK16890	Human brain expres
36	69.8	1.4	153	22	AAK42661	Human bone marrow
37	69.8	1.4	153	22	AAI23415	Probe #13348 for g
38	69.8	1.4	153	22	AAI48735	Probe #17421 used
39	69.8	1.4	153	22	AAI09041	Probe #9032 used t
40	69.8	1.4	153	23	ABA42283	Human liver single
41	69.8	1.4	153	24	AB816713	Human genome-deriv
42	65	1.3	65	24	ABNS4404	Mouse spliced tran
43	63	1.3	354	23	ABLI2699	Drosophila melanog
44	55.4	1.1	2043	18	AAI79627	Human Doc2-beta ge
45	54.4	1.1	369	22	ABA46385	Human breast cell

ALIGNMENTS

RESULT 1

AAF86684
ID AAF86684 standard; cDNA; 4980 BP.

AC AAF86684;

DT 25-JUL-2001 (first entry)

XX Mouse Rim2-encoding cDNA.

Mouse; murine; Rim2; Rim isoform; Rab3-interacting molecule;
Rab3-dependent intracellular vesicle fusion regulator;
GDP/GTP exchange factor; cAMP-GFII; neuron; endocrine tissue;
exocytosis; hormone secretion; neurotransmitter release;
presynaptic membrane; neuropathy; central nervous system disorder;
brain disorder; secretion disorder; pituitary gland; hypothalamus;
pancreatic beta-cell; diabetes; parotid gland; zinc finger; PDZ domain;
C2 domain; ss.

XX Mus musculus.

XX Key Location/Qualifiers

FT CDS 32..4804

FT /*tag= a

FT /product= "Mouse Rim 2"

FT /note= "Rim 2 is a novel isoform of Rim"

XX EP1090986-A1.

PD 11-APR-2001.

XX

```

31-JUL-2000; 2000EP-0116148.
08-OCT-1999; 99JP-0288372.
(JCRP-) JCR PHARM CO LTD.
(SEIN/) SEINO S.
Seino S, Shibasaki T, Ozaki N;
WPI; 2001-275168/29.
P-PSDB; AAB73488.
New Rim 2 proteins and nucleic acids encoding the protein useful in
diagnosing endocrine-related diseases or neuropathy and in developing
agents for the prevention or treatment of such diseases -
Claim 4; Page 19-31; 42pp; English.
This sequence represents cDNA encoding mouse Rim2. Rim2 is a novel
isoform of Rim (Rab3-interacting molecule, also referred to in the
specification as Rim1), and both proteins contain a zinc finger,
PDZ and two C2 domains. Rim2 is able to specifically interact with
the GTP/GDP exchange factor cAMP-GEFII, a cAMP sensor molecule, and
is believed to be a regulator of Rab3-dependent intracellular vesicle
fusion in neurons and endocrine tissue. It is therefore likely to
play a key role in neurotransmitter release at presynaptic membranes,
and in hormonal secretion. The Rim2 protein is useful in diagnosing
endocrine-related diseases or neuropathy and in developing agents for
the prevention or treatment of such diseases. Rim2-specific
hybridisation probes or antibodies are useful in the diagnosis of brain
or central nervous system disorders, or for diagnosis of secretion
disorders of the pituitary gland, hypothalamus, pancreatic beta-cells
and parotid gland.
Sequence 4980 BP; 1503 A; 1161 C; 1282 G; 1034 T; 0 other;
Query Match 100.0%; Score 4980; DB 22; Length 4980;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4980; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCTTCCCTAGGGTGGTTCGGCTCCGCCAACAATGTCGGCTCCGCTCGGGCCCCGGGGCGG 60
DB 1 GCTTCCCTAGGGTGGTTCGGCTCCGCCAACAATGTCGGCTCCGCTCGGGCCCCGGGGCGG 60
QY 61 CCGCGCTCCACCCCGGGCGCTCTCAACTCTCTCGCGAGCCGAGATCGCGGACTCAG 120
DB 61 CCGCGCTCCACCCCGGGCGCTCTCAACTCTCTCGCGAGCCGAGATCGCGGACTCAG 120
QY 121 CCACTCTACGAGAGGAGAGAAATCATCTGGCTGTTCATGATCTGTCAAGAAAGA 180
DB 121 CCACCTCTACGAGAGGAGAGAAATCATCTGGCTGTTCATGATCTGTCAAGAAAGA 180
QY 181 AGAGGAGAGGAGGAGTCCGTCTCAGATCAAGAGAGACCAAGACACACCCACACA 240
DB 181 AGAGGAGAGGAGGAGTCCGTCTCAGATCAAGAGAGACCAAGACACACCCACACA 240
QY 241 GTGGTTTCCCTTTAGTGGGATCACTGAACTGGTAAATAACGTTCTCGAGCCCCAGAAA 300
DB 241 GTGGTTTCCCTTTAGTGGGATCACTGAACTGGTAAATAACGTTCTCGAGCCCCAGAAA 300
QY 301 ACAACCCCAATGAGAGGAGGCCCGCCAGACAAGAGCTGCACCAACAAATTTGATAAGA 360
DB 301 ACAACCCCAATGAGAGGAGGCCCGCCAGACAAGAGCTGCACCAACAAATTTGATAAGA 360
QY 361 GCAAGTCAAGAGATGGGAGAGGAAATCGAGCGAGCGAGACAGAGAGGGTGATGCCCC 420
DB 361 GCAAGTCAAGAGATGGGAGAGGAAATCGAGCGAGCGAGACAGAGAGGGTGATGCCCC 420
QY 421 GACCTGTGGCATCTGCCACAAGACAATAATTGCAGATGGATGGCGGCATAATTTGTTCTTA 480
DB 421 GACCTGTGGCATCTGCCACAAGACAATAATTGCAGATGGATGGCGGCATAATTTGTTCTTA 480
QY 481 TTGCGAAACCAAGTTCGTGTCTCGATCTGGAGGTCGAGTGTCTTTACGCTCAACCAAGT 540

```

Db 1561 GCAGAGAAATGGAACCATGTTAAGGAATGATTTCTTTGAGTTCAGACCGAGTCCGAGTC 1620
Qy 1621 AGTGGGCGCCGCCACCAAGGCTCATTAATCCAGAGAAAGAGGTAAATGCGCCAGGT 1680
Db 1621 AGTGGGCGCCGCCACCAAGGCTCATTAATCCAGAGAAAGAGGTAAATGCGCCAGGT 1680
Qy 1681 TTCCTGAGCAGCTCCGAGAGAGAGCTGGCATCCACACCTGAGTATACAGAGCTGTGATCA 1740
Db 1681 TTCCTGAGCAGCTCCGAGAGAGAGCTGGCATCCACACCTGAGTATACAGAGCTGTGATCA 1740
Qy 1741 TGTGGAGCTGGAAGAGAGAGTGTGAGTGTGAGAGAGAGAGAGAGTCAAAAGGAGAAAGAA 1800
Db 1741 TGTGGAGCTGGAAGAGAGAGTGTGAGTGTGAGAGAGAGAGAGTCAAAAGGAGAAAGAA 1800
Qy 1801 AACTAGTGGAGAGAGAGTGTGAGTGTGAGAGAGAGAGAGTCAAAAGGAGAAAGAG 1860
Db 1801 AACTAGTGGAGAGAGAGTGTGAGTGTGAGAGAGAGAGAGTCAAAAGGAGAAAGAG 1860
Qy 1861 GATGTACTATGTTGGCCACTCTTTGGAAGAGAGTGTGAGTGTGAGAGAGAGAGTCAAA 1920
Db 1861 GATGTACTATGTTGGCCACTCTTTGGAAGAGAGTGTGAGTGTGAGAGAGAGAGTCAAA 1920
Qy 1921 GGACTCTGGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1980
Db 1921 GGACTCTGGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1980
Qy 1981 TAAGCACTCTGAGCTGGAGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 2040
Db 1981 TAAGCACTCTGAGCTGGAGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 2040
Qy 2041 ATTAATTAAGCTTTAAAGATGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 2100
Db 2041 ATTAATTAAGCTTTAAAGATGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 2100
Qy 2101 AAAGTGTGAGAGAGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 2160
Db 2101 AAAGTGTGAGAGAGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 2160
Qy 2161 AAAAAAGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 2220
Db 2161 AAAAAAGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 2220
Qy 2221 ATGGATGGAGCTTTTCAAGAGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 2280
Db 2221 ATGGATGGAGCTTTTCAAGAGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 2280
Qy 2281 ATCCAAACCTGAGACCAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 2340
Db 2281 ATCCAAACCTGAGACCAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 2340
Qy 2341 AATACCTGATGACGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 2400
Db 2341 AATACCTGATGACGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 2400
Qy 2401 AATGAGAGCTCTCTTATATCGTTTACCTTCAAGAGAGTGTGAGTGTGAGTGTGAGTGT 2460
Db 2401 AATGAGAGCTCTCTTATATCGTTTACCTTCAAGAGAGTGTGAGTGTGAGTGTGAGTGT 2460
Qy 2461 CCGCAGTCTTATCTGAGCAGCTTTCAATATAAATGAGTGTGAGTGTGAGTGTGAGTGT 2520
Db 2461 CCGCAGTCTTATCTGAGCAGCTTTCAATATAAATGAGTGTGAGTGTGAGTGTGAGTGT 2520
Qy 2521 GTTGATGTGATCAATTTTGGAGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 2580
Db 2521 GTTGATGTGATCAATTTTGGAGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 2580
Qy 2581 GAATCTTATGTATGATTTTCTTCCAGATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 2640
Db 2581 GAATCTTATGTATGATTTTCTTCCAGATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 2640
Qy 2641 AAAAAAGTCAAGAAAACTTTTGAACCCAAATGGAACCAAGTGTGAGTGTGAGTGTGAGTGT 2700
Db 2641 AAAAAAGTCAAGAAAACTTTTGAACCCAAATGGAACCAAGTGTGAGTGTGAGTGTGAGTGT 2700

Qy 2701 CCAACGAGAGAAATTCCTGGAACGAATCTGGAATAATACCTTTGGATCAAGCTAGAGT 2760
Db 2701 CCAACGAGAGAAATTCCTGGAACGAATCTGGAATAATACCTTTGGATCAAGCTAGAGT 2760
Qy 2761 TCGAGAGAGAGAGAGAGTTCCTTGAAGAGAGTTCCTTGAATTTGAATTTGAATTTGCT 2820
Db 2761 TCGAGAGAGAGAGAGAGTTCCTTGAAGAGAGTTCCTTGAATTTGAATTTGAATTTGCT 2820
Qy 2821 AGATGATGAGCGCTACTGATATAAGCTGAGACCAATGATGTCCTCTCAATTCGCACTCC 2880
Db 2821 AGATGATGAGCGCTACTGATATAAGCTGAGACCAATGATGTCCTCTCAATTCGCACTCC 2880
Qy 2881 TCGCCCTTCCCATATCTGCCCGGAGGAGCTCCATCTGAGAGAGCCCAACGCGAGGCT 2940
Db 2881 TCGCCCTTCCCATATCTGCCCGGAGGAGCTCCATCTGAGAGAGCCCAACGCGAGGCT 2940
Qy 2941 GCAAGGTCGAGAGAGATTAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 3000
Db 2941 GCAAGGTCGAGAGAGATTAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 3000
Qy 3001 GGGAGTGTGTCAGATTAATGACACAAATGCGCGGAGTCTTCAAGCTCCAGTTCGCT 3060
Db 3001 GGGAGTGTGTCAGATTAATGACACAAATGCGCGGAGTCTTCAAGCTCCAGTTCGCT 3060
Qy 3061 GCCAGAACAGTGTCAATCAATCAATTTGCTCACCATCAGGCTCTCTCATCGAGTAGA 3120
Db 3061 GCCAGAACAGTGTCAATCAATCAATTTGCTCACCATCAGGCTCTCTCATCGAGTAGA 3120
Qy 3121 TGTATATGAGAGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 3180
Db 3121 TGTATATGAGAGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 3180
Qy 3181 ACAGGGGACCCGAGGAGACAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3240
Db 3181 ACAGGGGACCCGAGGAGACAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3240
Qy 3241 ACACGCTGTCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 3300
Db 3241 ACACGCTGTCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 3300
Qy 3301 TAGAGAGCCATATCAAGATTCAGATCAAGAACAGAGAACAGAGAACAGAGAACAGAGAAC 3360
Db 3301 TAGAGAGCCATATCAAGATTCAGATCAAGAACAGAGAACAGAGAACAGAGAACAGAGAAC 3360
Qy 3361 CACCGCTCCAGATCTCTGAAAGCTCTGATACAAACCTCATGAGTGTGAGTGTGAGTGT 3420
Db 3361 CACCGCTCCAGATCTCTGAAAGCTCTGATACAAACCTCATGAGTGTGAGTGTGAGTGT 3420
Qy 3421 AATGATGGAAGATCTGCCCTCTCTTCACTGCTCTTATCGAGTGTCTCACCTCTGCTACCG 3480
Db 3421 AATGATGGAAGATCTGCCCTCTCTTCACTGCTCTTATCGAGTGTCTCACCTCTGCTACCG 3480
Qy 3481 GTCTGTCCAGACAGCCCATCAAGTATCTCCGGAACAGAGACAGAGAGAGAGAGAGAGAG 3540
Db 3481 GTCTGTCCAGACAGCCCATCAAGTATCTCCGGAACAGAGACAGAGAGAGAGAGAGAGAG 3540
Qy 3541 ACAGCTTCCACAAAGGAGACATTTGGAGAGAGTGTGATGATATGAGAGAGAGAGAGAG 3600
Db 3541 ACAGCTTCCACAAAGGAGACATTTGGAGAGAGTGTGATGATATGAGAGAGAGAGAGAG 3600
Qy 3601 CCAATGAAATCTTAACAAATACAAAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 3660
Db 3601 CCAATGAAATCTTAACAAATACAAAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 3660
Qy 3661 TTACATTCGAGTGTGCTCAGATGCTCAGATGCTCAGATGCTCAGATGCTCAGATGCTCAG 3720
Db 3661 TTACATTCGAGTGTGCTCAGATGCTCAGATGCTCAGATGCTCAGATGCTCAGATGCTCAG 3720
Qy 3721 TAAATCTCGAGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3780
Db 3721 TAAATCTCGAGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3780

Db	1021	TTTACAGATCATCGTCTGGGAGATTATGGCCGCGATCGATCAAAATCTTTATGGAG	1080
Qy	4638	TGGCCGAGATCTCTTAGATGAATCTGGAACTATCCAAATGATGGATGGATGTTCAAC	4697
Db	1081	TGGCCGAGATCTCTTTTAGATGAATCTAGAGCTATCCAAATGATGGATGGATGTTCAAC	1140
Qy	4698	TCCTCCCTCCTCTCCCTCTAGTAGATCCAACTTGGCCACTCTGACAGAGAGCTTCCC	4757
Db	1141	TTTTCCCACTTCTCTCTCTAGTAGATCCAACTTGGCCCTCTGACAGAGAGCTTCCC	1200
Qy	4758	AATCGTCTCTGGAAATCTTACCGGACCTTCTTACTCTCTGTTTCATAGCAACT---ATAAA	4814
Db	1201	AATCATCTCTGGAAATCTTCAACTGGACCTCTTACTTCTCTGTTTCATAGCAACTGTA	1260
Qy	4815	ACGTGTTGTACACACACACACACATACAAAACCGAAGAAA	4856
Db	1261	ATTGTTGTACACACACACACACATACAAAACCGAAGAAA	1302
RESULT 3			
AAI98097			
ID	AAI98097 standard; cDNA; 6637 BP.		
XX	AAI98097;		
DT	04-DEC-2001 (first entry)		
XX	Human neuroblastoma expressed polynucleotide SEQ ID NO 40.		
DE	Human; neuroblastoma; ss.		
XX	Homo sapiens.		
OS	WO200166733-A1.		
XX	13-SEP-2001.		
XX	02-MAR-2001; 2001WO-JP01631.		
XX	07-MAR-2000; 2000JP-0159195.		
FR	12-MAY-2000; 2000JP-0140387.		
XX	(CHIB-) CHIBA PREPECTURE.		
PA	(HISM) HISAMITSU PHARM CO LTD.		
XX	Makagawara A;		
PI	WPI; 2001-602630/68.		
XX	Nucleic acids for prognosis of human neuroblastoma comprise nucleic acids expressed by human neuroblastomas -		
PT	Claim 1; Page 97-101; 159pp; Japanese.		
XX	The invention relates to nucleic acids (AAI98058-AAI98161) or their homologues expressed by human neuroblastomas useful for detecting genes expressed by neuroblastoma and for analysing their structure and function. The nucleic acids are useful for the diagnosis and prognosis of neuroblastoma.		
CC	Sequence 6637 BP; 2020 A; 1423 C; 1511 G; 1683 T; 0 other;		
CC	Query Match 14.4%; Score 715; DB 22; Length 6637;		
CC	Best Local Similarity 57.2%; Pred. No. 1.2e-194;		
CC	Matches 1771; Conservative 0; Mismatches 1045; Indels 279; Gaps 15;		
Qy	89	CTCTCTCCGACCGCGAGATCGGACCTTACGCACTTCCGCAATGATGGATGTTCAATC	148
Db	366	CTCTCTCCGACCGCGAGATCGGACCTTACGCACTTCCGCAATGATGGATGTTCAATC	425
Qy	149	ATCTCGGCTGTCTATCGATCGTCTCAGAGAAAGAGAGAGAGCGATCTCTCTCTCAAG	208

[illegible]

QY	1271	GC	AAAGCTGTAACTAGAGATTC	CGAGATTCCTCTCTAAGATGATGACCAATCAAGG	1330
DB	1506	CC	CGCACCGAGCGCGCGC	-----CGCGCCTGCGGAGCGCAGGCGCGCAAA	1553
QY	1331	CAA	AGATCTGTATCTGAACTG	TAGAGCTGCAATGCAAAAACCAACGATCGTATTCAAATGCGAA	1390
DB	1554	CG	CGCGCGCGCGCGCAGCGAG	GGCCTCGCGCGCGGACTCGCGCGCGGCTTACTCGGCTGAG	1613
QY	1391	AG	AACCTCGAGAGGCTCAGG	GACAAAGTTCCTTATTCACAAAGGACCTCAAAATCATAGTCTCT	1450
DB	1614	AGA	ACTGCGGAGACACAGG	CGCGCGCGCGCGCTAA-----CGAACCAACAGCCCG	1667
QY	1451	CC	CACCCCTCGCGGAGCCCT	TATACCGCTGTGTATAGACCAACGATGAGGCGCGCTACTCC	1510
DB	1668	CG	CGCGCGCCAGACATGG	CGCGCGGTTCCCGCAGAGCCCCCGAGCTCAAGGCCGAGGAGCCC	1727
QY	1511	CT	ACGGAACAGCACACAT	TTAGATCCAGCTCTGCTGTGAGGAAAAACGACCGGAAAAA	1570
DB	1728	CT	CAGGAGCAGAGCGCC	CTGACCCCGACCTCGGC-----1762	
QY	1571	AT	GAAACCAATGTTAAG	GAATGATTCCTTGAGTTTACAGCAAGCTCCGAGTCAGTGAAGCCG	1630
DB	1763	---	GGTCTCTATGCTGCG	GAACGATCTTTTGTAGCTACAGCCAGTCCGATGTCGCTCGCGCGC	1820
QY	1631	CC	CCCAACCAAGGCTCAT	ATAATCCAAAGAAAGGAGGTAAATAATGCGCCAGGTTTCACTGAGC	1690
DB	1821	TC	CCCGCCCAAGCGCA	CCGGTCCAGAGAGGCGCGCAGAGCGCGCATGTTCGTGAGC	1880
QY	1691	AG	CTCGGAGAGAGCTGG	CATCCACACTGTAGTATACAGCTGTGATGATGATGAGTGAAGCTG	1750
DB	1881	AG	CTCTGAGGAGGAGCG	CTGTGACGCCCGAGTACACAGCTGCGAGGACGTGAGCTG	1940
QY	1751	GA	AGCGAGAGCTGTAG	TGAGAAAGGGGACAGTCAAAAGGGAAAAAGAAAACCTAGCTGAG	1810
DB	1941	GAG	AGCGAGAGCGT	CAAGCGAAGAGTGATTT-----1972	
QY	1811	CAG	GAGTTTTGTGCGGAT	TCAACACACAGGTTGAGAGACAAAGAAAAAGATGTACTAT	1870
DB	1973	-----	-----1972		
QY	1871	GG	TGGCCACTCTTTGG	AAGAGATTTGGAAATGCTCTGACCTCAGATTAAGGACTCTGGG	1930
DB	1973	-----	GGATTAATCTGG	TGCGATCC-----1993	
QY	1931	GT	AGATACCTGTAGT	AGCACAACCTTAAACGAGAGCATAGCCATAGTCATGATAGCACCTT	1990
DB	1994	-----	TGCC	ACGTGGCACAGCGCGGAGACATCACTATTAGTTCGATCTCT	2039
QY	1991	GT	CACCTTGGCAGCCAT	CCAAAGATGAGATTCGCTTAATTTGCTCGTATTTTATTAATAAG	2050
DB	2040	GT	AACCTGGCAACCAT	CTAAGAGGGGGACCGGATTAATTTGACGCTGTATTCTCTTAACAG	2099
QY	2051	CG	TTTAAAAAGATGG	CGAGTGTACTCTCAGAGCTCAGGAGCAATGCTGGCGCTTAAAGGTGTGA	2110
DB	2100	AGA	CAACCA-----	TGCCCAAGACTCAGGTGCATTTCTGGGCTCGAAGATGTT	2150
QY	2111	GG	AGAAAGATGATGA	ATCAGTTCGACTTTGTGCAATTTATTAACAAAGTAAAAAAGGA	2170
DB	2151	GG	AGGAAAAATGACT	GTAGACACTTTGTGCTTTTCAACAACAAAGTAAAGAGGGT	2210
QY	2171	AG	TTTAGCTGATACT	GTAGACATCTTACACAGCGTGTATGAGTCTGTGAATGCAATGGG	2230
DB	2211	AG	CTAGCAGATGTAG	TTGACACTTAAGGACGCGGATGAGTCTTAGATGGAATGGT	2270
QY	2231	AG	CTATTTCAGAGGCC	CAATTTGAGGAAGTTTACACATTTCTTAGATTCCAACCT	2290
DB	2271	AA	ACCCCTCGCGGAG	CTACAAATGAAGAAGTTTACACATTTTGAATCAATCA	2330
QY	2291	GA	ACCAAGATTGAG	CTGTTGTTTTCACGCCAATTTGAGATATTCCTAGATACCTGAT	2350
DB	2331	GA	ACCTCAAGATTGA	AAATTTGTTTTCACGCCCTATTGGTGACATTCGCCGATTCCTGAG	2390

QY 2351 AGCAGCATGACACAGTGGATTCAGTCTCATTTGATCTCAAAAATGAGCCGT 2410
Db |||||
QY 2391 AGCTCCACCCCTCCAGTGGAGTCCAGTTCAGTTCCTTTGATCTCAGAGATGGAAGG 2450
Db |||||
QY 2411 CTTCTATATCCGTACCTCACCAGTGGCTGGGATGCTGAGGATGTCGCCAGTTC 2470
Db |||||
QY 2451 CTTCCATTTCTGTTATTTCTCAGAGTCTGGAGCTTAAAGATGCCCAAGTC 2510
Db |||||
QY 2471 TTATCTGACAGCTTTCAATAAATATGTTTTCAGAGGTTGTCACAGTTCATGAT 2530
Db |||||
QY 2511 TTACAGAGGCACTTTCTGAGTGTGTATGATAAAGTGGACACAGCTGATGTA 2570
Db |||||
QY 2531 ACATTTTGGAGCAAGATCTCCCTCCAGGAGATGGGAGGCAAGATCTTAT 2590
Db |||||
QY 2571 AATGTTCTGCAAGCAACAGATCTACCTGCTAGTAGTAGAGCTCTCGAATCCCTAT 2630
Db |||||
QY 2591 GTTAGATTTACTTCTCCAGATAGAGTGTATAAATAAGAGAGACAAAACGTC 2650
Db |||||
QY 2631 GTAAATGTATTTCTTCAGATAGAGTGTATAAATAAGAGAGGACCAACAGTA 2690
Db |||||
QY 2651 AAGAAATCTTGGAAACCCAAATGGACACAGATCTTCAATTTATCTCTGTCACCGAAG 2710
Db |||||
QY 2691 AAGAAATCTAGAACCAAAATGGATCAAACTTTGCTATTCATGATGATGTA 2750
Db |||||
QY 2711 GAATTCGGTGAAGATGCTGGAAATACCTTTGGGATCAAGCTAGATTCGAGAGAA 2770
Db |||||
QY 2751 GAATTTAGAGAACGATGTTAGAAATACTGTGTGGGACCAACCAAGAGTGCAGAGAA 2810
Db |||||
QY 2771 GAGAGCAATCTTAGGAGAGATTTTAAATGGAATGGAACAGCTTTGCTAGATGATG 2830
Db |||||
QY 2811 GAAGTGAATTTCTTGGAGAGATCTCTAGAGATTTGGAGACAGCGCTTTAGATGAA 2870
Db |||||
QY 2831 CGGCTGTGTATAGTGCAGACCCATGATGCTCTCTCATTTGCCACTCCCTCGCCCTCC 2890
Db |||||
QY 2871 CGGCTGTGTATAGTGCAGACCCATGATGCTCTCTCATTTGCCACTCCCTCGCCCTCC 2930
Db |||||
QY 2891 CCATATCTGCCCGAGGAGCTCCATGAGAGAGCCCAACGGCAGGCTGCAGAGGTCG 2950
Db |||||
QY 2931 CTTTCATGSCCAAGGCGCATATTCATGAGAGAAAGCTCTAGCAAAAAGCTACAAAGATCT 2990
Db |||||
QY 2951 AAGAGATTAAGTGCAGATGCTGACTAGCTGCGAGATGGCTGGGCTAGTGT 3010
Db |||||
QY 2991 CAGCGAATCAGTGTATGATGATCTCAGATTTAGAGTTGATGTTGTTATGGCTAGT 3050
Db |||||
QY 3011 TC-----AGATTATCGACACATGCGCGCATCTTCAAAGCTCCACGTTGTCGGTGCCA 3064
Db |||||
QY 3051 CTTCCAGTAGGCTATAGTCTAGTGTAGAGAAAGTAAATCTACACATTAATCTGCGCA 3110
Db |||||
QY 3065 GAACAGTCAATGTCATCAATCAATGCTCAGATGAGGCTCTCTCATGAGTATGAT 3124
Db |||||
QY 3111 GAACAGCAAGAACAACTCATACGCTCAGCTTCAGTATCTCTCATGCGGCAATGAT 3170
Db |||||
QY 3125 ATAGAGAGCAAGATCATGCTGCTAGTGGCC 3159
Db |||||
QY 3171 CAGGGAAGCCGCTTACGTTTACAAATGTGCC 3205
Db |||||

RESULT 4
ID AAS94817
XX AAS94817 standard; DNA; 7325 BP.
AC AAS94817;
XX
DT 14-FEB-2002 (first entry)
XX
DE Human DNA sequence #72 expressed during foam cell differentiation.
KW Human; foam cell differentiation; atherosclerosis; cerebral stroke;
KW cardiovascular disorder; coronary artery disease; gene therapy; ds.
OS Homo sapiens.
XX
FN WO200177389-A2.

18-OCT-2001.
04-APR-2001; 2001WO-US11128.
05-APR-2000; 2000US-195106P.
(INCV-) INCVTE GENOMICS INC.
Shiffman D, Somogyi R, Lawn R, Seilhamer JJ, Porter GJ, Mikita T;
Tai J;
WPI; 2002-010925/01.
Composition useful for diagnosis of conditions, disorders or diseases
associated with atherosclerosis, comprises several polynucleotides that
are differentially expressed in foam cell development -
Claim 1; Page 123-125; 315pp; English.
The present invention relates to the isolation of human polynucleotide
sequences that are differentially expressed during foam cell
differentiation. The polynucleotide sequences of the invention or a
composition comprising these polynucleotides are useful as a high
throughput method for detecting altered expression of one or more
polynucleotides in a sample. The polynucleotides can be used in the
diagnosis of disorders associated with foam cell development such as
atherosclerosis, cerebral stroke, and cardiovascular disorders such as
coronary artery disease. The polynucleotide sequences can also be used
as PCR primers and probes. The polynucleotide sequences of the invention are also
useful in gene therapy. AAS94746-AAS95021 represent the human
polynucleotide sequences of the invention which are differentially
expressed during foam cell differentiation.

Query Match 8.0%; Score 398.2; DB 24; Length 7325;
Best Local Similarity 67.9%; Pred. No. 3.4e-103;
Matches 556; Conservative 0; Mismatches 263; Indels 0; Gaps 0;
QY 3977 GCGCCCTTGGTACCACTGGGCAAGAGCGGCGATCTAGCATTTGGGCGCAAAATGCTAGT 4036
Db |||||
QY 4037 ATTGTTGCTCTCAGGAAAGTCCGAGTCTCTCACTCAGCCAAACCGAGAGAGA 4096
Db |||||
QY 641 ATCTGGGCGCTGACTCAGTGGGCAAGAGCGACATCCAGCTTCGCGACCTGAGGGGCC 700
QY 4097 GGTAAAAAGCTACGAGGACATCTTTCAGAGAGACGAGAGACCGGCTAGCAGTGGAGATG 4156
Db |||||
QY 4157 AGGAACTGGATGACCCGCGAGGCGAGGATCCAGATGCGAGATGCGAGCTAT 4216
Db |||||
QY 761 CGAGCGCGGCTCAGACGCGAGCGGCGAGGCTCCACCGATGGGAGCAGCAGCAAC 820
QY 4217 AGCTCGGAGCAAAATCTGATCTTCCCTGGGCTGGGCTCTGAGCAGGCTTACT 4276
Db |||||
QY 821 AGCTCGGAGCGCACTTCTTCCACTTCCACTTCCACTTCCACTTCCACTTCCACTTCCAGC 880
QY 4277 GATTTCTGATGCTGGGCGCTGCTCAGCTAGTGGGAGCGGCGAGCCAGACCTGGGCTACTCT 4336
Db |||||
QY 881 GATTTCTGATGCTGGGCGCTGCTCAGCTAGTGGGAGCGGCGAGCAGCAGCAGCAGCAGCA 940
QY 4337 GCAATGGTGCATTCAGGTTGGGATGATGATGATGATGATGATGATGATGATGATGATGAT 4396
Db |||||
QY 941 CCCATGGAGATGTCATTCGATTCATTCGATTCATTCGATTCATTCGATTCATTCGATTCGAT 1000
QY 4397 ATCCGGGCGGCGGCTTGTGTTAAACAGGCTTCAGAGACATGCGCAGCAGCAGTATGTC 4456
Db |||||
QY 1001 ATTGAAGCTGGGCGCTGAGCCCCCAACAGGCTCCAAATCCCTCCAGGCACTATATC 1060
QY 4457 AAGGTGTATCTGTTAGACAAACGAGTCTGCTATAGCCAAAGAAACCAAGGTCGGGAGA 4516

[illegible][illegible]

RESULT 6

ABAI1415
ID ABAI1415 standard; cDNA; 526 BP.
XX
AC ABAI1415;
XX
DT 23-JAN-2002 (first entry)
XX
DE Human nervous system related polynucleotide SEQ ID NO 422.
XX
KW Human; nontropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antineoplastic; anti-HIV; antibacterial; vulnary;
KW antiparkinsonian; antiskinning; antianemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ss.
XX
OS Homo sapiens.
XX
FN WO200159063-A2.
XX
PD 16-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01334.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 23-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235835.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.

PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250391.
 PR 01-DEC-2000; 2000US-0251160.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Rosen CA, Barash SC, Ruben SM;
 XX P-PSDB; ABB15089.
 DR WPI; 2001-541565/60.
 XX P-PSDB; ABB15089.
 XX
 PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating nervous system
 PT cancers and metastases -
 XX
 PS Claim 1; SEQ ID NO 422; 1701pp + Sequence Listing; English.
 XX
 CC The invention relates to novel genes (AB11004-AB21534) and proteins
 CC (AB14678-AB18001) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC and ovarian cancer and other cancers of the adrenal gland, bone,
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 526 BP; 151 A; 134 C; 132 G; 103 T; 6 other;
 Query Match 6.4%; Score 318.8; DB 22; Length 526;
 Best Local Similarity 92.7%; Pred. No. 5.2e-81;
 Matches 332; Conservative 2; Mismatches 24; Indels 0; Gaps 0;
 OY 3231 GAATGATAGACACCGTTCATGATGACCATCTCTTCATAGACAGAGGATTG 3290
 DB 1 GAATGACACAGATCGTGTGATGATGACCATCTCTTCATAGACAGAGGATTG 60
 OY 3291 AAGCAGCAGATAGACAGCCATATCAGATCCAGATCAACAGAACGGCTCTCTCAG 3350
 DB 61 AAGCAGCAGATAGACAGCCATATCAGATCCAGATCAACAGAACGGCTCTCTCAG 120
 OY 3351 ACGGACACACCGCTCCAGATCTCTGACGCTCTGATCAAAACCTCATGAGGTGGA 3410
 DB 121 ACGGACACACCGCTCCAGATCTCTGACGCTCTGATCAACCTCATGAGGTGGA 180
 OY 3411 TGCCTTCATTAATGATGAGATGCGCCCTCTCTTACCTGCTTATGAGGTCTCACC 3470
 DB 181 TGCCTTCATTAATGATGAGATGCGCCCTCTCTTACCTGCTTATGAGGTCTCACC 240
 OY 3471 CTCGTACCGGCTCTGTCAGACAGCCATCAAGTACTCCGGACACGACGAGGGGCC 3530
 DB 241 CTCGTACTGGGCTCTGTCAGACAGCCATCAAGTACTCCAGTCGACGACGAGGGGCC 300

OY 3531 GACAGCTTCCACAGCTTCCACCAAGGGAACATTGGAGAGAGTCTATGANTATAGA 3588
 DB 301 GACAGCTTCCACAGCTTCCACCAAGGGAACATTGGAGAGAGTCTATGANTATAGA 358
 DB 301 GACAGCTTCCACAGCTTCCACCAAGGGAACATTGGAGAGAGTCTATGANTATAGA 358

RESULT 7

AAV90372/c

ID AAV90372 standard; cDNA; 355 BP.

XX AAV90372;

XX 15-FEB-1999 (first entry)

XX EST clone DMI18.

XX Human; secreted protein; expressed sequence tag; EST; haematopoiesis;
 KW tissue growth; activin; inhibin; chemotaxis; chemokinesis; haemostatic;
 KW receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;
 KW gene therapy; ss.

XX Homo sapiens.

XX W09845436-A2.

XX 15-OCT-1998.

XX 10-APR-1998; 98WO-US06955.

XX 10-APR-1997; 97US-0838821.

XX (GENY) GENETICS INST INC.

XX Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;

XX Racie LA, Spaulding V, Treacy M;

XX WPI; 1999-070077/06.

XX New polynucleotides encoding human secreted proteins - derived from
 PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
 PT ovary, pituitary, retina and colon cDNA libraries.

XX Claim 1; Page 519-520; 618pp; English.

XX The present sequence represents a human expressed sequence tag (EST).
 CC The polynucleotide, which is a secreted EST, and the encoded protein
 CC are predicted to have useful biological activities which would make
 CC them suitable for treating, preventing or ameliorating medical
 CC conditions in humans and animals, although no supporting data is
 CC given. Suggested activities include nutritional activity, immune
 CC stimulating or suppressing activity, haematopoiesis regulating
 CC activity, tissue growth activity, activin/inhibin activity,
 CC chemotactic/chemokinetic activity, haemostatic and thrombolytic
 CC activity, receptor/ligand activity, anti-inflammatory activity,
 CC cadherin/tumour invasion suppressor activity, tumour inhibition
 CC activity. The polynucleotide may also be useful for gene therapy.

XX Sequence 355 BP; 93 A; 82 C; 83 G; 97 T; 0 other;

Query Match 4.2%; Score 207.6; DB 20; Length 355;

Best Local Similarity 90.2%; Pred. No. 5.1e-49;

Matches 222; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

OY 3844 CAGTGTCTTTATCCTCAAAATGCAAAACAGACAGATGGCGTGTGGGAGACTTGC 3903

DB 247 CAGTGTCTTTATCCTCAAAATGCAAAACAGACAGATGGCGTGTGGGAGACTTGC 188

OY 3904 CAAAGCACCAGCATCAGTGGAGACATGTCTCTCTGAGAGATTCACGCGACGATC 3963

DB 187 AANAAGCACCAGCATCAGTGGAGACATGTCTCTCTGAGAGATTCACGCGACGATC 128

OY 3964 CGACATCTGAGTGGGCGCCCTGGGTACCGATGGGACAGCGCGATCTAGCATTTGGGCG 4023

169 AAAAAGCACCAGCATCATGTGGACACATGTCTCATCTGGACAGAAATGTGGCAGCGATC 238

3964 CGACACTGCAGTGGCGCCCTGGGTACCACTGGCCACAGACCGCGCATCTAGCATTTGGGC 4023

229 TGCACCTGCAGTGGCGACCTTTGGGCGCCAGTGGCGCAAAAAGCGCGCGCTCTAGCCTTGGTGC 288

4024 CAAAAATGGTAGCTATTGTGTGGTCTCTCAGCGAAAATCGCAGTGCGCTCTCACTCAGCCCA 4083

289 CAAAAATGGTAGCTATTGTGTGGTCTCTCAGCGAAAATCGCAGTGCGCTCTCTCAGCTCAGCGG 348

4084 AACCGA 4089

349 ACTCGA 354

RESULT 9

AAC74934

ID AAC74934 standard; cDNA; 378 BP.

XX

AC AAC74934;

XX

DT 08-FEB-2001 (first entry)

XX

XX Human ORFX ORF489 polynucleotide sequence SEQ ID NO:977.

XX

XX Human; open reading frame; ORFX; detection; cytosolic; hepatotropic; vulnary; antipsoaric; antiparkinsonian; nootropic; neuroprotective; anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant; immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antinflammatory; antiviral; antibacterial; antifungal; antirheumatic; antithyroid; antianemic; gene therapy; cancer; proliferative disorder; hypertension; neurodegenerative disease; osteoarthritis; graft-vs host disease; cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS; cholesterol ester storage; systemic lupus erythematosus; infection; severe combined immunodeficiency; malaria; autoimmune disorder; asthma; allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound; bone damage; cartilage damage; antinflammatory disease; coagulation; thrombosis; contraceptive; ss.

XX

OS Homo sapiens.

XX

XX WO200058473-A2.

XX

XX 05-OCT-2000.

XX

XX 31-MAR-2000; 2000WO-US08621.

XX

XX 31-MAR-1999; 99US-0127607.

XX

XX 02-APR-1999; 99US-0127636.

XX

XX 05-APR-1999; 99US-0127728.

XX

XX 30-MAR-2000; 2000US-0540763.

XX

XX (CURA-) CURAGEN CORP.

XX

XX Shimkets RA, Leach M;

XX

XX WPI; 2000-602362/57.

XX

XX P-PSDB; AAB40725.

XX

XX Novel nucleic acids and peptides derived from open reading frame X, useful for treating e.g. cancers, proliferative disorders,

XX

XX neurodegenerative disorders and cardiovascular disease -

XX

XX Claim 5; Page 956; 5507pp; English.

XX

XX AAC74446 to AAC7606 encode the proteins given in AAB40237 to AAB43397, which represent the human ORFX open reading frames 1 to 3161. The ORFX sequences have activities such as: cytosolic; hepatotropic; vulnary; antipsoaric; antiparkinsonian; nootropic; neuroprotective; osteopathic; anticonvulsant; antirheumatic; immunosuppressant; immunostimulant; cardiant; thrombolytic; coagulant; vasotonic;

XX DB Human breast cell single exon nucleic acid probe #4132.
 XX KW Human; microarray; single exon probe; gene expression; breast;
 XX OS disease; cancer; ss.
 XX OS Homo sapiens.
 XX PN WO200157271-A2.
 XX PD 09-AUG-2001.
 XX PF 30-JAN-2001; 2001WO-US00662.
 XX PR 04-FEB-2000; 2000US-0180312.
 XX PR 26-MAY-2000; 2000US-0207456.
 XX PR 30-JUN-2000; 2000US-0608408.
 XX PR 03-AUG-2000; 2000US-0632366.
 XX PR 21-SEP-2000; 2000US-0234687.
 XX PR 27-SEP-2000; 2000US-0236359.
 XX PR 04-OCT-2000; 2000GB-0024263.
 XX PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX DR WPI; 2001-496933/54.
 XX KW New spatially-addressable set of single exon nucleic acid probes,
 XX PT useful for measuring gene expression in sample derived from human
 XX PT breast, comprises number of single exon nucleic acid probes -
 XX PS Claim 1; SEQ ID NO 4132; 327pp + sequence listing; English.
 XX CC The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human breast and BT 474 cells. The method involves contacting
 CC the probes with a collection of detectably labelled nucleic acids
 CC derived from mRNA of human breast, and then measuring the label
 CC bound to each probe of the microarray. The probes are useful for
 CC verifying the expression of regions of genomic DNA predicted to
 CC encode proteins. They are useful for gene discovery, and for
 CC determining predisposition and/or prognosing breast disease. Gene
 CC expression analysis is useful for assessing the toxicity of chemical
 CC agents on cells. The microarray of this invention presents a far greater
 CC diversity of probes for measuring gene expression, with far less bias
 CC than expressed sequence tag microarrays. The method is suitable for
 CC rapid production of functional information from genomic sequence. The
 CC present sequence is a single exon nucleic acid probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX SQ Sequence 365 BP; 111 A; 62 C; 90 G; 102 T; 0 other;
 Query Match 2.3%; Score 113.8; DB 22; Length 365;
 Best Local Similarity 72.2%; Pred. No. 6.1e-22;
 Matches 148; Conservative 0; Mismatches 57; Indels 0; Gaps 0;
 QY 3837 GGAAATCAGTGTCTTTACATCCAAATGCAAAACAGACAGATGGCGGTGGGGAAGA 3896
 DB 161 GTAATATAGTTTCATTTACCCCAAAATGCAAGGACAGACGGATTCAGGAAGT 220
 QY 3897 ACTTGACCAAAAGCACCAGCATCAGTGGAGACATGTCTCACTGGAGAGATGACGGCA 3956
 DB 221 CCATCATGAGAGACACCAAGTGTCTGAGAGATGTACACATGGAGCATATATGACGGCA 280
 QY 3957 GCCAGTCCGACACTGCACTGGCGCCCTGGGTACAGTGGCAGAGAGCGGATCTAGCA 4016
 DB 281 GCCAGTCCGACACTGCGGTGTACAGTGGAGCAGGTGGAGAGAACCGGATCCAGCC 340
 QY 4017 TTGGGGCCAAATGCTAGTATTGT 4041

DB 341 TTAGTCCAAAGTGGTGGCCATAGT 365
 RESULT 12
 ABA55935
 ID ABA55935 standard; DNA; 365 BP.
 XX AC ABA55935;
 XX DT 01-FEB-2002 (first entry)
 XX DE Human foetal liver single exon nucleic acid probe #4240.
 XX KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
 XX OS Homo sapiens.
 XX PN WO200157277-A2.
 XX PD 09-AUG-2001.
 XX PF 30-JAN-2001; 2001WO-US00669.
 XX PR 04-FEB-2000; 2000US-0180312.
 XX PR 26-MAY-2000; 2000US-0207456.
 XX PR 30-JUN-2000; 2000US-0608408.
 XX PR 03-AUG-2000; 2000US-0632366.
 XX PR 21-SEP-2000; 2000US-0234687.
 XX PR 27-SEP-2000; 2000US-0236359.
 XX PR 04-OCT-2000; 2000GB-0024263.
 XX PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX DR WPI; 2001-483447/52.
 XX KW Human genome-derived single exon nucleic acid probes useful for
 XX PT analyzing gene expression in human fetal liver -
 XX PS Claim 1; SEQ ID NO 4240; 639pp + sequence listing; English.
 XX CC The invention relates to a single exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human foetal
 CC liver. The single exon nucleic acid probes may be used for predicting,
 CC measuring and displaying gene expression in samples derived from human
 CC foetal liver. The present sequence is a single exon nucleic acid
 CC probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX SQ Sequence 365 BP; 111 A; 62 C; 90 G; 102 T; 0 other;
 Query Match 2.3%; Score 113.8; DB 22; Length 365;
 Best Local Similarity 72.2%; Pred. No. 6.1e-22;
 Matches 148; Conservative 0; Mismatches 57; Indels 0; Gaps 0;
 QY 3837 GGAAATCAGTGTCTTTACATCCAAATGCAAAACAGACAGATGGCGGTGGGGAAGA 3896
 DB 161 GTAATATAGTTTCATTTACCCCAAAATGCAAGGACAGACGGATTCAGGAAGT 220
 QY 3897 ACTTGACCAAAAGCACCAGCATCAGTGGAGACATGTCTCACTGGAGAGATGACGGCA 3956
 DB 221 CCATCATGAGAGACACCAAGTGTCTGAGAGATGTACACATGGAGCATATATGACGGCA 280
 QY 3957 GCCAGTCCGACACTGCACTGGCGCCCTGGGTACAGTGGCAGAGAGCGGATCTAGCA 4016
 DB 281 GCCAGTCCGACACTGCGGTGTACAGTGGAGCAGGTGGAGAGAACCGGATCCAGCC 340
 QY 4017 TTGGGGCCAAATGCTAGTATTGT 4041
 DB 341 TTAGTCCAAAGTGGTGGCCATAGT 365

4017 TTGGGCCCCAAAATGCTAGCTATTCT 4041
341 TTAGTGCCAAAAGTGGTTGCCATAGT 365

RESULT 14
AAK04148
ID AAK04148 standard; DNA; 365 BP.
XX AC AAK04148;
XX AC AAK04148;
XX 05-NOV-2001 (first entry)
XX DE Human brain expressed single exon probe SEQ ID NO: 4139.
XX KW Human; brain expressed exon; gene expression analysis; probe;
XX KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
XX KW epilepsy; cancer; ss.
XX OS Homo sapiens.
XX XX WO200157275-A2.
XX PN 09-AUG-2001.
XX PD 30-JAN-2001; 2001WO-US006667.
XX PP 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-48899/53.
XX Single exon nucleic acid probes for analyzing gene expression in human
XX hearts -
XX Claim 1; SEQ ID No 4068; 530pp; English.
XX The present invention relates to single exon nucleic acid probes for
XX measuring human gene expression in a sample derived from human heart. The
XX present sequence is one such probe. The probes may be used for
XX predicting, measuring and displaying gene expression in samples derived
XX from the human heart via microarrays. By measuring gene expression, the
XX probes are useful for predicting, diagnosing, grading, staging,
XX monitoring and prognosing diseases of the human heart and vascular system
XX e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
XX congenital heart disease.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 365 BP; 111 A; 62 C; 90 G; 102 T; 0 other;
XX

Query Match 2.3%; Score 113.8; DB 22; Length 365;
Best Local Similarity 72.2%; Pred. No. 6.1e-22;
Matches 148; Conservative 0; Mismatches 57; Indels 0; Gaps 0;
3837 GCAAAATCAGTGTCTTTACATCCAAAATGCAAAACAGACAGATGGCGGTGTCGGGGAAGA 3896
161 GTAATATAGTTCATTTACCCCAAAATGCAAGGACAGCGATTCAGGAGAT 220
3897 ACTTGACCAAAAGCAACAGCATCAGTGAGACATGTGCTCACTGGGAGAGATGACGCA 3956
221 CCATCATGAAGACGACACAGTGTCACTGGAGAGATGTACACACTGGAGCATAATGACGCA 280
3957 GCCAGTCCGACATCTGAGTGGCGCCCTGGGTACCATGTCGACAGAGCGGATCTAGCA 4016
281 GCCAGTCCGACATCTGAGTGGCGCCCTGGGTACCATGTCGACAGAGCGGATCTAGCA 4016

Query Match 2.3%; Score 113.8; DB 22; Length 365;
Best Local Similarity 72.2%; Pred. No. 6.1e-22;
Matches 148; Conservative 0; Mismatches 57; Indels 0; Gaps 0;
3837 GCAAAATCAGTGTCTTTACATCCAAAATGCAAAACAGACAGATGGCGGTGTCGGGGAAGA 3896
161 GTAATATAGTTCATTTACCCCAAAATGCAAGGACAGCGATTCAGGAGAT 220
3897 ACTTGACCAAAAGCAACAGCATCAGTGAGACATGTGCTCACTGGGAGAGATGACGCA 3956
221 CCATCATGAAGACGACACAGTGTCACTGGAGAGATGTACACACTGGAGCATAATGACGCA 280
3957 GCCAGTCCGACATCTGAGTGGCGCCCTGGGTACCATGTCGACAGAGCGGATCTAGCA 4016
281 GCCAGTCCGACATCTGAGTGGCGCCCTGGGTACCATGTCGACAGAGCGGATCTAGCA 4016

4017 TTGGGCCCCAAAATGCTAGCTATTCT 4041
341 TTAGTGCCAAAAGTGGTTGCCATAGT 365

RESULT 13
ABA25602
ID ABA25602 standard; DNA; 365 BP.
XX AC ABA25602;
XX AC ABA25602;
XX 23-JAN-2002 (first entry)
XX DE Probe #4068 for gene expression analysis in human heart cell sample.
XX KW Human; gene expression; heart; microarray; vascular system; probe;
XX KW cardiovascular disease; hypertension; cardiac arrhythmia;
XX KW congenital heart disease; ss.
XX OS Homo sapiens.
XX XX WO200157274-A2.
XX PN 09-AUG-2001.
XX PD 30-JAN-2001; 2001WO-US006666.
XX PP 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-48899/53.
XX Single exon nucleic acid probes for analyzing gene expression in human
XX hearts -
XX Claim 1; SEQ ID No 4068; 530pp; English.
XX The present invention relates to single exon nucleic acid probes for
XX measuring human gene expression in a sample derived from human heart. The
XX present sequence is one such probe. The probes may be used for
XX predicting, measuring and displaying gene expression in samples derived
XX from the human heart via microarrays. By measuring gene expression, the
XX probes are useful for predicting, diagnosing, grading, staging,
XX monitoring and prognosing diseases of the human heart and vascular system
XX e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
XX congenital heart disease.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 365 BP; 111 A; 62 C; 90 G; 102 T; 0 other;
XX

Query Match 2.3%; Score 113.8; DB 22; Length 365;
Best Local Similarity 72.2%; Pred. No. 6.1e-22;
Matches 148; Conservative 0; Mismatches 57; Indels 0; Gaps 0;
3837 GCAAAATCAGTGTCTTTACATCCAAAATGCAAAACAGACAGATGGCGGTGTCGGGGAAGA 3896
161 GTAATATAGTTCATTTACCCCAAAATGCAAGGACAGCGATTCAGGAGAT 220
3897 ACTTGACCAAAAGCAACAGCATCAGTGAGACATGTGCTCACTGGGAGAGATGACGCA 3956
221 CCATCATGAAGACGACACAGTGTCACTGGAGAGATGTACACACTGGAGCATAATGACGCA 280
3957 GCCAGTCCGACATCTGAGTGGCGCCCTGGGTACCATGTCGACAGAGCGGATCTAGCA 4016
281 GCCAGTCCGACATCTGAGTGGCGCCCTGGGTACCATGTCGACAGAGCGGATCTAGCA 4016

Search completed: December 11, 2003, 15:51:13
Job time : 1181 secs

```
QY 4017 TTGGGGCCAAATGGTAGCTATTGT 4041
Db 341 TTAGTCCCAAGTGGTTGCCATAGT 365

RESULT 15
AAK29630
ID AAK29630 standard; DNA; 365 BP.
XX
AC AAK29630;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed single exon probe SEQ ID NO: 4187.
XX
KW Human; bone marrow expressed exon; gene expression analysis; probe;
XX microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
OS Homo sapiens.
XX
PN WO200157276-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00668.
XX
PR 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488900/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -
XX
PS Example 4; SEQ ID NO: 4187; 658pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC the probes of the invention.
XX
SQ Sequence 365 BP; 111 A; 62 C; 90 G; 102 T; 0 other;

Query Match 2.3%; Score 113.8; DB 22; Length 365;
Best Local Similarity 72.2%; Pred. No. 6.1e-22;
Matches 148; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 3837 GCAAAATCAGTGTCTTTACATCCAAATGCAAAACAGACAGATGGCGGTGCGGGGAAGA 3896
Db 161 GTAAATATAGTTCATTACCCCAAAATGCAAGGAGACGATGGGACTTCAGGNAGAT 220

QY 3897 ACTTGACCAAAAGCACCGACATCAGTGGAGACATGTCTCACTGGAGAGAGATGACGGCA 3956
Db 221 CCATCATGAAGAGCACCAGTGTCACTGGAGAGATGTACACACTGGAGCATATGAGGCA 280

QY 3957 GCCAGTCCGACACTGAGTGGCGGCTGGGTACAGTGGCAAGAGCGGCGATCTAGCA 4016
Db 281 GCCAGTCCGACACACTGTGGGTACAGTGGAGCAGGTGGAAAGAAACCGAGATCCAGCC 340

QY 4017 TTGGGGCCCAAAATGGTAGCTATTGT 4041
Db 341 TTAGTCCCAAGTGGTTGCCATAGT 365
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 11, 2003, 15:31:19 ; Search time 306 Seconds
(without alignments)
7183.294 Million cell updates/sec

Title: US-09-617-099b-2

Perfect score: 4980

Sequence: 1 gttccctagggtggttcgg.....acattgtccctagcaagg 4980

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgm2_6/prodata/2/ina/5A COMB.seq.*
- 2: /cgm2_6/prodata/2/ina/5B COMB.seq.*
- 3: /cgm2_6/prodata/2/ina/6A COMB.seq.*
- 4: /cgm2_6/prodata/2/ina/6B COMB.seq.*
- 5: /cgm2_6/prodata/2/ina/PCUTUS COMB.seq.*
- 6: /cgm2_6/prodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	67.2	1.3	7218	1	US-08-232-463-14
2	47.6	1.0	2244	3	US-09-094-714A-48
3	47.6	1.0	2245	4	US-09-225-743-24
4	46	0.9	533	6	5482709-5
5	45	0.9	3489	2	US-08-728-323A-1
6	45	0.9	3489	4	US-09-298-568-1
7	45	0.9	3489	4	US-09-410-399-1
C 8	45	0.9	32207	2	US-08-770-379-20
C 9	45	0.9	32207	3	US-08-757-662A-20
C 10	45	0.9	32207	4	US-09-230-371A-20
C 11	43.8	0.9	1037	4	US-09-181-585-3
C 12	43.8	0.9	1159	4	US-09-181-585-1
C 13	43.8	0.9	1471	4	US-09-181-585-2
C 14	43.4	0.9	397	3	US-09-253-691-3
15	42.4	0.9	4466	4	US-09-410-551B-20
16	42.4	0.9	4478	4	US-09-410-551B-16
17	42.4	0.9	4547	4	US-09-410-551B-22
18	42.4	0.9	4571	4	US-09-410-551B-18
19	42.4	0.9	77536	4	US-09-410-551B-1
20	42	0.8	203	3	US-09-043-303-7
21	42	0.8	1071	4	US-09-397-787-159
C 22	41.4	0.8	51259	3	US-08-781-891-209
C 23	41.4	0.8	51259	4	US-09-618-166-209
24	41.2	0.8	975	4	US-09-328-352-3332
25	41.2	0.8	6558	4	US-09-491-356C-7
26	41.2	0.8	6794	4	US-09-491-356C-2
27	40.8	0.8	477	3	US-09-135-994-1

28	40.8	0.8	477	4	US-09-684-843A-1
29	40.8	0.8	543	6	5273901-6
C 30	40.8	0.8	3891	3	US-09-036-315-1
C 31	40.6	0.8	16442	3	US-08-781-891-208
C 32	40.6	0.8	16442	4	US-09-618-166-208
C 33	40.2	0.8	9603	4	US-09-203-895-1
34	39.6	0.8	154	1	US-08-469-802B-6
35	39.6	0.8	154	2	US-08-267-803B-5
36	39.6	0.8	165	3	US-09-043-303-17
37	39.6	0.8	168	1	US-08-469-802B-4
38	39.6	0.8	168	2	US-08-267-803B-4
39	39.6	0.8	171	1	US-08-469-802B-5
40	39.6	0.8	171	2	US-08-267-803B-5
41	39.6	0.8	195	1	US-08-469-802B-2
42	39.6	0.8	195	2	US-08-267-803B-2
43	39.6	0.8	234	1	US-08-469-802B-3
44	39.6	0.8	234	2	US-08-267-803B-3
45	39.6	0.8	3376	1	US-08-320-559-29

ALIGNMENTS

RESULT 1

US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 INMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:

LENGTH: 7218 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

CLONE: PTZgpt-F1s

US-08-232-463-14

Query Match

1.3%; Score 67.2; DB 1; Length 7218;


```

Db      334 GAGCAG 339

RESULT 5
US-08-728-323A-1
; Sequence 1, Application US/08728323A
; Patent No. 5948676
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: Immediate Early Protein From Kaposi's
; TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA
; TITLE OF INVENTION: Encoding Same And Uses Thereof
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,323A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 5575/52268/JPW/MSC/SKS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-278-0400
; TELEFAX: 212-391-0525
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3489 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3489
US-08-728-323A-1

Query Match      0.9%; Score 45; DB 2; Length 3489
Best Local Similarity 48.6%; Pred. NO. 0.013;
Matches 123; Conservative 0; Mismatches 130; Indels

QY      169 TCAGAGAGAAAGACGAGGAGGAGCAGTCCCTGCTCAAGATCAAGAGAG
Db      1989 TGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGGATGAG
QY      229 ACAACCGACACAGTGGTTCCTTTTGTGGGATCTCTGACTGTGTAATAT
Db      2049 TGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAG
QY      289 GCCTCAGCAAAAACAACTCATGTGAGAGGAGCCCTCAGCAAAAGCTGCACT
Db      2109 GGATGAGCAGGAGCAGCAGGATGAGCAGCAGGATGAGCAGCAGGATGAGCAG
QY      349 AATGTATAAGGACCACTCAAGAGAGATGGGAGGGAATCTCGCAGCAGC
Db      2169 GCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAG
QY      409 GGGTGTATGCCCG 421

```


Query Match	0.9%	Score 43.8;	DB 4;	Length 1159;
Best Local Similarity	47.0%;	Pred. No. 0.013;		
Matches 135;	Conservative 0;	Mismatches 152;	Indels 0;	Gaps 0;

QY	123	ACCTCAGGAGAGAGAGAAATCATCTCTGCTGTTCATGATCGTCAGAGAAGAAG	182
Db	766	ACATCAATAGTAAATAATAGATAATATATTTTAAAAATGCAGCAGCAGCAGCAGC	707
QY	183	AGGAGAGAGGAGAGTCGTGCTCAGATCAAAGAGACACAAAGACAAACCGACACAGT	242
Db	706	AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC	647

RESULT 14
US-09-253-691-3
; Sequence 3, Application US/09253691
; Patent No. 6124100
; GENERAL INFORMATION:
; APPLICANT: Dong Kyu JIN
; TITLE OF INVENTION: Diagnostic Method and Kit for Neuropsychiatric Diseases
; TITLE OF INVENTION: Using Trinucleotide Repeats Sequence
; FILE REFERENCE: 1942/36
; CURRENT APPLICATION NUMBER: US/09/253,691
; CURRENT FILING DATE: 1999-02-22
; EARLIER APPLICATION NUMBER: KR 98-6,278
; EARLIER FILING DATE: 1996-02-26

NUMBER OF SEQ ID NOS: 3
SOFTWARE: WordPerfect 6.1/Windows
SEQ ID NO 3
LENGTH: 397
TYPE: DNA
ORGANISM: human
US-09-253-691-3

Query Match 0.9%; Score 43.4; DB 3; Length 397;
Best Local Similarity 53.9%; Pred. No. 0.0079;
Matches 89; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 245 TTTCCTTTAGTGGGATCACTGATGTTAATAAGCTTTCAGCCCCCGAAGAAACAA 304
DB 98 TTCACTTTTGAATGTTTCAGACAGCAGCAAAAGCAGCAACAGCAGCAGCAGCAG 157
QY 305 CCCAATGAGAGAGAGCCCGCAGCAAAAGCTGCACCAACCAATTTGAAATGTATAAGGAGCA 364
DB 158 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 217
QY 365 GTCAAGAGAGATGGGAGAGGAATCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 409
DB 218 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 262

RESULT 15

US-09-410-551B-20
Sequence 20, Application US/09410551B
Patent No. 6503737
GENERAL INFORMATION:
APPLICANT: KOSAN BIOSCIENCES, Inc.
APPLICANT: REEVES, CHRISTOPHER
APPLICANT: CHU, DANIEL
APPLICANT: KHOSLA, CHAITAN
APPLICANT: SANTI, DANIEL
APPLICANT: WU, KAI
TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA
TITLE OF INVENTION: CONSTRUCTS THEREFOR
FILE REFERENCE: 30062-20026.00
CURRENT APPLICATION NUMBER: US/09/410.551B
CURRENT FILING DATE: 1999-10-01
PRIOR APPLICATION NUMBER: US 60/139,650
PRIOR FILING DATE: 1999-06-17
PRIOR APPLICATION NUMBER: US 60/123,810
PRIOR FILING DATE: 1999-03-11
PRIOR APPLICATION NUMBER: US 60/102,748
PRIOR FILING DATE: 1998-10-02
NUMBER OF SEQ ID NOS: 72
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 20
LENGTH: 4466
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: DNA encoding synthetic
OTHER INFORMATION: PKS synthase fragment
FEATURE:
NAME/KEY: CDS
LOCATION: (9)..(4454)
US-09-410-551B-20

Query Match 0.9%; Score 42.4; DB 4; Length 4466;
Best Local Similarity 50.5%; Pred. No. 0.094;
Matches 103; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 3330 CAGAACAGCGGCTCTCTAGAGGGGACCAACCCGCTCCAGATCTCTGAAGTCTTG 3389
DB 272 CGGACCCGCGCCAGCGCGCGGCGGACGAGCGCGCTGGCGATCGTGGGAATGGCTG 331
QY 3390 ATACAACTCTATGAGGTGCGATGCTTCAATTAATGACCTGGAAGATCGCCCTCTTCAC 3449
DB 332 CCGGCTGCCCGGGGTGGGTACCCGAGAGCTGTGGCACTCTGTGGCATCGGCAC 391

QY 3450 CTGCCTTATCGAGGTCTCACTCTGTATCGGCTCTGTCTCCAGCAGCCCATCTAGTCTC 3509
DB 392 CGACCCCATCACGGAGTTCCCGACGCGACCGCGGCTGGGACCTTCAGCGGATCTACGACC 451
QY 3510 CCGGAACAGGACGAGAGGGGCGGAC 3533
DB 452 GGACCCCGACCGGATCGGCAAGAC 475

Search completed: December 11, 2003, 23:16:04
Job time : 309 secs

Db 2341 AATACCTGATAGCAGCATGCAAACTGGAATCCAGTCTAGCTCAITTTGAATCTCAAA 2400
Qy 2401 AATGGAAGCTCTCTATATCCGTATACCTCAACCAAGATCTCTGGGATGCTGAGGATGT 2460
Db 2401 AATGGAAGCTCTCTATATCCGTATACCTCAACCAAGATCTCTGGGATGCTGAGGATGT 2460
Qy 2461 CCCGAGTCTCTATCTGAGCAGCTCTCAATAAATACTATGTTTCAAGGTTGCTCAACA 2520
Db 2461 CCCGAGTCTCTATCTGAGCAGCTCTCAATAAATACTATGTTTCAAGGTTGCTCAACA 2520
Qy 2521 GTTGATAGTATCAATTTTGGAGCAAGGATCTCCCTTCCAGGAGAGATGGGAGCCAG 2580
Db 2521 GTTGATAGTATCAATTTTGGAGCAAGGATCTCCCTTCCAGGAGAGATGGGAGCCAG 2580
Qy 2581 GAATCCTATGTTAAGATTTATCTCTCCAGATAGAGTGAATAAATAAGAGAGAC 2640
Db 2581 GAATCCTATGTTAAGATTTATCTCTCCAGATAGAGTGAATAAATAAGAGAGAC 2640
Qy 2641 AAAAAAGTCAAGAAACTTTGGAACTTGGAACTTGGAACTTGGAACTTGGAACTTGG 2700
Db 2641 AAAAAAGTCAAGAAACTTTGGAACTTGGAACTTGGAACTTGGAACTTGGAACTTGG 2700
Qy 2701 CCACGAGAGAAATTCGGTGAACGAATGCTGGAAATTTACCTTTGGGATCAAGCTAGGT 2760
Db 2701 CCACGAGAGAAATTCGGTGAACGAATGCTGGAAATTTACCTTTGGGATCAAGCTAGGT 2760
Qy 2761 TCCAGAGAGAGAGAGCAATTTCTAGAGAGATTTTAAATGAAATGGAACAGCTTTGCT 2820
Db 2761 TCCAGAGAGAGAGAGCAATTTCTAGAGAGATTTTAAATGAAATGGAACAGCTTTGCT 2820
Qy 2821 AGATGATGAGCCGACATGATAGCTGAGAGCCAGATGATGATGATGATGATGATGATG 2880
Db 2821 AGATGATGAGCCGACATGATAGCTGAGAGCCAGATGATGATGATGATGATGATGATG 2880
Qy 2881 TCGCCCTTCCCATATCTGCCCGAGGAGCTCCATGAGAGAGCCCAACGCGAGGCT 2940
Db 2881 TCGCCCTTCCCATATCTGCCCGAGGAGCTCCATGAGAGAGCCCAACGCGAGGCT 2940
Qy 2941 GCAGAGTCCAGAGAGATAGTGAAGTGTCTGATCACTGAGAGAGAGAGAGAGAGAG 3000
Db 2941 GCAGAGTCCAGAGAGATAGTGAAGTGTCTGATCACTGAGAGAGAGAGAGAGAGAG 3000
Qy 3001 GCGAGTGTCTGAGATATTCGACACAAATGCGCGGATCTTCAAGCTCCAGCTTTGCTGT 3060
Db 3001 GCGAGTGTCTGAGATATTCGACACAAATGCGCGGATCTTCAAGCTCCAGCTTTGCTGT 3060
Qy 3061 GCAGAGCAAGTCAATGTCAATCAATCTCTCAATCAAGGCTCTCTCATCGAGTGA 3120
Db 3061 GCAGAGCAAGTCAATGTCAATCAATCTCTCAATCAAGGCTCTCTCATCGAGTGA 3120
Qy 3121 TGTATAGGAGAGCAAGGTCTGCTGCTAGTGGCCCTCTCTCTCAAGGATGTGA 3180
Db 3121 TGTATAGGAGAGCAAGGTCTGCTGCTAGTGGCCCTCTCTCTCAAGGATGTGA 3180
Qy 3181 ACAGGAGCAGGAGGAGCAGTCTACTGCGCATTAACACAAATTAGCCGATGATAG 3240
Db 3181 ACAGGAGCAGGAGGAGCAGTCTACTGCGCATTAACACAAATTAGCCGATGATAG 3240
Qy 3241 ACAACCGTGTCAATGATGACCACTACTCTTCAGATAGAGCAGGATTTGAGAGCAGCA 3300
Db 3241 ACAACCGTGTCAATGATGACCACTACTCTTCAGATAGAGCAGGATTTGAGAGCAGCA 3300
Qy 3301 TAGACAGCCATATCAAGATCCAGATCAACAGAAACAGGCTCTCTCTAGAGAGGAGCCAC 3360
Db 3301 TAGACAGCCATATCAAGATCCAGATCAACAGAAACAGGCTCTCTCTAGAGAGGAGCCAC 3360
Qy 3361 CACCCCTCCAGATCTCTGAAAGCTCTGATCAAACTCTCATGAGGTGATGCTTTCAAT 3420
Db 3361 CACCCCTCCAGATCTCTGAAAGCTCTGATCAAACTCTCATGAGGTGATGCTTTCAAT 3420
Qy 3421 AATGACTGGAAGATCTGCCCTCTCTTCACTGCTTTATCGAGGTCTCACCCCTGACCGG 3480
Db 3421 AATGACTGGAAGATCTGCCCTCTCTTCACTGCTTTATCGAGGTCTCACCCCTGACCGG 3480

Qy 3481 GTCTGTCCAGCAAGCCCATCAAGTACTCCGGGACAGACGAGGGGCGGACAGCTTCC 3540
Db 3481 GTCTGTCCAGCAAGCCCATCAAGTACTCCGGGACAGACGAGGGGCGGACAGCTTCC 3540
Qy 3541 ACAGCTTCCACCAAGGAGAACTTGGAGAGAGTGTCTATGATATAGAGAGAGAAATCG 3600
Db 3541 ACAGCTTCCACCAAGGAGAACTTGGAGAGAGTGTCTATGATATAGAGAGAGAAATCG 3600
Qy 3601 CCAATGAAACTTAAACAAATACAAACAGAGTAGCGGATCTAGACCTTCAGATCGAGAGA 3660
Db 3601 CCAATGAAACTTAAACAAATACAAACAGAGTAGCGGATCTAGACCTTCAGATCGAGAGA 3660
Qy 3661 TTACCAATTCGAGATATCGCTCAGAGTAGGATCCACATAGAGGGCGAGATCTGTTCCAC 3720
Db 3661 TTACCAATTCGAGATATCGCTCAGAGTAGGATCCACATAGAGGGCGAGATCTGTTCCAC 3720
Qy 3721 TAAATCTCCGACAGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3780
Db 3721 TAAATCTCCGACAGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3780
Qy 3781 TCGTTTCAGCAGCAAGCTTACATGTCCTGATCAATCAAGAGGGCGGATGAGAGAGAGAA 3840
Db 3781 TCGTTTCAGCAGCAAGCTTACATGTCCTGATCAATCAAGAGGGCGGATGAGAGAGAGAA 3840
Qy 3841 AATCAGTGTCTTTTATCTCAAAATGCAAAACAGACAGATGGGCTGTCCGGAGAGAACTT 3900
Db 3841 AATCAGTGTCTTTTATCTCAAAATGCAAAACAGACAGATGGGCTGTCCGGAGAGAACTT 3900
Qy 3901 GACCAAAAGCAGCAGATCTGATGAGAGATGATGATGATGATGATGATGATGATGATGAT 3960
Db 3901 GACCAAAAGCAGCAGATCTGATGAGAGATGATGATGATGATGATGATGATGATGATGAT 3960
Qy 3961 GTCCGACACTGCAAGTGGGCGCTTGGGATACAGAGGAGAGAGAGAGAGAGAGAGAGAG 4020
Db 3961 GTCCGACACTGCAAGTGGGCGCTTGGGATACAGAGGAGAGAGAGAGAGAGAGAGAGAG 4020
Qy 4021 GGCAGAAATGCTAGTGTATGTTGCTCTCAACGAGAAAGTCCAGTGTCTCTCACTCAG 4080
Db 4021 GGCAGAAATGCTAGTGTATGTTGCTCTCAACGAGAAAGTCCAGTGTCTCTCACTCAG 4080
Qy 4081 CCAACCGAGAGAGAGAGTAAAGCTACGAGCACTGTTCTAGAGAGAGAGAGAGAGAGAG 4140
Db 4081 CCAACCGAGAGAGAGAGTAAAGCTACGAGCACTGTTCTAGAGAGAGAGAGAGAGAGAG 4140
Qy 4141 GCTAGCAGTGGAGATGAG 4200
Db 4141 GCTAGCAGTGGAGATGAG 4200
Qy 4201 CAGCATGAAACAGCTATAGCTCCGAGAGAAATCTGATCTTCTGCGGCTCCGCTGCGCTC 4260
Db 4201 CAGCATGAAACAGCTATAGCTCCGAGAGAAATCTGATCTTCTGCGGCTCCGCTGCGCTC 4260
Qy 4261 TGAAGCAGCTTCAAGTGTATCTGATGATGATGATGATGATGATGATGATGATGATGAT 4320
Db 4261 TGAAGCAGCTTCAAGTGTATCTGATGATGATGATGATGATGATGATGATGATGATGAT 4320
Qy 4321 GACCTTGGCTTACTCTGCAATGCGGTGATCTGAGTGGAGATGATGATGATGATGATGAT 4380
Db 4321 GACCTTGGCTTACTCTGCAATGCGGTGATCTGAGTGGAGATGATGATGATGATGATGAT 4380
Qy 4381 GCTGAGGTAGAGAAATCATCCCGGCGCTTGTGTGTAACCAAGCTTCTCAAGCACT 4440
Db 4381 GCTGAGGTAGAGAAATCATCCCGGCGCTTGTGTGTAACCAAGCTTCTCAAGCACT 4440
Qy 4441 GCGAGCAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4500
Db 4441 GCGAGCAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4500
Qy 4501 AACCAAGTGGGAG 4560
Db 4501 AACCAAGTGGGAG 4560

1391 AGAATCGAGGCTCAGGACAAAGTTCTTATCCACAAAGGACCTCAATCATATGTCCT 1450
1614 AGAATCGAGGCTCAGGACAAAGTTCTTATCCACAAAGGACCTCAATCATATGTCCT 1667
1451 CCACCCCTCGGCGGACCTTATACCGCTTTGATAGACCAAGCATGAGCGCGCTGACTCC 1510
1668 CCGCGCGCCAGACATGCGCGGTTCCCGCAGAAAGCCCGGAGCTCAAGCCCGAGGCCC 1727
1511 CTAGGAAACAGCACCACTTAGATCCAGCTCTGCTGTGAGGAAAAACGAAGCCGAGAAAA 1570
1728 CTAGGAAACAGCACCACTTAGATCCAGCTCTGCTGTGAGGAAAAACGAAGCCGAGAAAA 1762
1571 ATGGAACACCATGTTTAAAGGAATGATCTTTCAGTTTTCAGACCAAGTCCGAGTCAAGTCCGCG 1630
1763 --GGTCTCATGCTTGCAGAACGACTCTTTGAGTCTCAGACCAAGTCCGAGTCAAGTCCGCG 1820
1631 CCCCCCAAGCGCTCATTAATCCAGAAAGGAGGTAAATGCGCCAGAGTTTCACTGAGC 1690
1821 TCCCCCGCCCAAGCGCTCCAGAGAGGCGGCGAGAGCGCGAGATGTCGCTGAGC 1880
1691 AGCTCGAGGAGGAGCTGGCATCCACACTGATATACAGCTGTGATGATGAGGAGCTG 1750
1881 AGCTCTGAGGAGGAGGCGGTGTCGACCGCCGAGTACACAGCTGCGAGGAGCTGAGCTG 1940
1751 GAAAGCGAGAGTCTGAGTCAAGAAAGGAGGAGCTCAAGAGGAGGAGGAGGAGGAGGAGGAG 1810
1941 GAGAGCGAGAGCTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1972
1811 CAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1870
1973 ----- 1972
1871 GGTGGCCACTCTTTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1930
1973 -----GGAATTATTAAGTGGATCC----- 1993
1931 GTAGATACCTGTAGTAGCAACCTTTAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2039
1994 -----TGCCAGCTGGGACAGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2050
1991 GTGACCTGGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2050
2040 GTAACTGGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2099
2051 CTTTAAAGAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2110
2100 AGAACCAACCA-----TGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2150
2111 GGAG 2170
2151 GGAG 2210
2171 AGTTAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2230
2211 AGCTTAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2270
2231 AGCTTAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2290
2271 AAACCCCTCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2330
2291 GAAACCAAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2350
2331 GAACTCAAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2390
2351 AGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2410
2391 AGCTCCACCTCGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2450
2411 CTTCTATATCGTTTACCTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2470
2451 CTTCTATATCGTTTACCTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2510
2471 TTATCTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2530

2511 TTACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2570
2531 ACAATTTTGGAGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2590
2571 AATGTTCTGAGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2630
2591 GTTAAGATTTACTTCTTCCAGATAGAGTGAATAAATAAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 2650
2631 GTAAATAATGATTTTCTTCCAGATAGAGTGAATAAATAAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 2690
2651 AGGAAATCTTTGAGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2710
2691 AGGAAATCTTTGAGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2750
2711 GAATTCGAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2770
2751 GAATTTAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2810
2771 GAGAG 2830
2811 GAAAGTGAATTTCTTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2870
2831 CCGCACTGATTAAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2890
2871 CCGCAATGATTAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2930
2891 CCATATCTGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2950
2931 CTTTATGAG 2990
2951 AAGAGATTAAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3010
2991 CAGCAATCAAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3050
3011 TC-----AGATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3064
3051 CTTCCAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3110
3065 GAACAGGATCAATGATCAATCAATGATCAATGATCAATGATCAATGATCAATGATCAATGATCAATGATCAAT 3124
3111 GAACAGGAG 3170
3125 ATAGAG 3159
3171 CAGGAG 3205

RESULT 3

US-10-240-965-72

; Sequence 72, Application US/10240965

; Publication No. US20030165924A1

; GENERAL INFORMATION:

; APPLICANT: INCYTE GENOMICS, INC.

; APPLICANT: SHIFFMAN, Dov

; APPLICANT: SOMOGYI, Roland

; APPLICANT: LAWE, Richard M.

; APPLICANT: SEILHAMER, Jeffrey J.

; APPLICANT: PORTER, Gordon J.

; APPLICANT: MIKITA, Thomas

; APPLICANT: TAL, Julie

; TITLE OF INVENTION: GENES EXPRESSED IN FOAM CELL DIFFERENTIATION

; FILE REFERENCE: PA-0025 PCT

; CURRENT APPLICATION NUMBER: US/10/240,965

; CURRENT FILING DATE: 2002-10-04

; PRIOR APPLICATION NUMBER: 60/195,106

; NUMBER OF SEQ ID NOS: 276

; SOFTWARE: PERL Program

; SEQ ID NO 72

; LENGTH: 7325

; TYPE: DNA

; ORGANISM: Homo sapiens

```

; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030165924A1 977757.3
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 4103, 4411, 4430
; OTHER INFORMATION: a, t, c, g, or other
US-10-240-965-72

Query Match      8.0%; Score 398.2; DB 13; Length 7325;
Best Local Similarity 67.9%; Pred. No. 1.2e-110;
Matches 556; Conservative 0; Mismatches 263; Indels 0; Gaps 0;

QY 3977 GGCGCCCTGGGTACCAAGAGCGCGGATCTAGCAATTTGGGCCCCAAATGGTAGCT 4036
DB 581 GGCGCTGGGACCAACACCGCCCAAGAGCGCGGAGCGCTGGTGGTGGGCGC 640
QY 4037 ATTGTTGGTCTCTACGGAAGAGTGGCAGTGTCTCTCACTAGCCAAACCGAGGAGA 4096
DB 641 ATCGTGGGCTGACTCAGTGGAGCAAGGACACACTCCAGCTTCCGCGAGCTGAGGGGCC 700
QY 4097 GGTAAAAAGCTACGAGACCTGTTTCAGAGAGAGCAGGAGCGCGGCTAGCAGTGGAGATG 4156
DB 701 ACCAGAGCTTGGGAGCAACATCCGCGGAGCAGGAGACAGCATCGCGGTGGAGATG 760
QY 4157 AGAACTGGATGACCCCGCAGGCGGAGCGGGAATCCACAGATGGCAGCATGAACAGCTAT 4216
DB 761 CGGAGCGGGTCAACCGCCAGGCGGAGCGCGGAGTCCAGCATGGGAGCCTCAACAGAAC 820
QY 4217 AGCTCGAGAGGAATCTGATCTTCCCTGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 4276
DB 821 AGCTCGAGCGGACGCTGATCTTCTTCCCACTACCCCGGTAGGGGCTGAAAGCCAGTTTCAG 880
QY 4277 GATTCTCTGATGGCTGGGCGCTGCTCAGCTAGTGGGAGCGGAGCGGCTGGCTCTCTCTCT 4336
DB 881 GATTCTCTGATGGCTGGGCGCTGAGCAGCTCAGATGTTGGGCGGAGCAGACACTGGCAACCA 940
QY 4337 GCATGGGTGACATTCAGGTGGGAATGATGATGATGATGATGATGATGATGATGATGATG 4396
DB 941 CCATGGGAGATGTCACATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATG 1000
QY 4397 ATCGGCGGCGGCGGCGCTTGTGTAACCGAGTTCAGACACTGCGCAGCAGCGTATGTC 4456
DB 1001 ATTGAAGCTCGGCGGCTGACCCCAACCGAGGCTCCAAATCCCTCCCGAGCCTATATC 1060
QY 4457 AAGCTGTATCTGTAGACAGGAGTCTGCATAGCCCAAGAAAGAAAGAAAGAGGAGGAGG 4516
DB 1061 AAGTTTACTCTGTGAGATGGGCGCTCTTGGCCAGAGAGAGAGAGAGAGAGAGAGAGAG 1120
QY 4517 AAGACCTGGAGCCCTGTACACGAGCTCTTGTCTTGGAGAGGAGGAGGAGGAGGAGG 4576
DB 1121 AAGACCTGTGATCCCTGTACACGAGGCTCTGCTCTTTGACGAGGAGGAGGAGGAGG 1180
QY 4577 GTGTTACAGATCATCTCTGGGAGATGATGCTGATGATGATGATGATGATGATGATGATG 4636
DB 1181 GTCTCGAGGTGATCTCTGGGAGACTATGGCGGATGATGATGATGATGATGATGATGATG 1240
QY 4637 GTGCGCCAGATCTCTTAGATGAATGGAATCAATCAATGATGATGATGATGATGATGATG 4696
DB 1241 ATGGCCAGATCATCTGAGAGCTGTGACCTCAGGCGCGCGGTACCGGCTGGTACAAA 1300
QY 4697 CTCTTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4756
DB 1301 CTCTTCCCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1360
QY 4757 CAATCTCTCTGGAAGTCTTACCGACCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4795
DB 1361 CAGTCTCTCTGAGAGTGCACACGAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1399

```

RESULT 4
US-09-864-761-4068
; Sequence 4068, Application US/09864761

```

; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn. Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecmics-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 4068
; LENGTH: 365
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL035633.15
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.9
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.6
; US-09-864-761-4068

Query Match      2.3%; Score 113.8; DB 9; Length 365;
Best Local Similarity 72.2%; Pred. No. 8.5e-24;
Matches 148; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 3837 GGAAATCAGTCTCTTTACATCCAAATATCCAAACAGCAGATGGCGCTGTCTGGGGAAGA 3896

```

Db 161 GTAATATAGTTCATTATCCCAAAATGCAAGCAGCGATGGGACTTCAGAGAT 220
QY 3897 ACTTGACAAAGACACAGCATCATGTGAGACATGTCTCATCTGAGAGAAATGACGGCA 3956
Db 221 CCATCATGAAGAGACACCATGTCTAGTGGAGAGATGTATACACTGGAGCATATGACGGCA 280
QY 3957 GCCAGTCCGACACATGAGTGGCGCCCTGGGTACCATGGTGGCAGAGAGCGGATCTAGCA 4016
Db 281 GCCAGTCCGACACATGAGTGGGTACATGGTGGAGAGTGGAGAGAAACCGAGATCCAGCC 340
QY 4017 TTGGGGCCAAATGTTAGCTATTCT 4041
Db 341 TTAGTCCAAAGTGGTTGCCATCT 365

RESULT 5

US-10-029-386-3580
; Sequence 3580, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Ranz, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: ABOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 3580
; LENGTH: 568
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL035633.18
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.81
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1
; OTHER INFORMATION: NT HIT: G416306469, EVALUATE 1.00e-111
; OTHER INFORMATION: EST HUMAN HIT: BF951698.1, EVALUATE 0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: Q03188, EVALUATE 7.10e+00
US-10-029-386-3580

Query Match 2.1%; Score 105.8; DB 13; Length 568;
Best Local Similarity 74.0%; Pred. No. 3.5e-21;
Matches 134; Conservative 0; Mismatches 47; Indels 0; Gaps 0;
QY 4624 ATCTTTATGGAGTGGCCAGATATCTTTAGATGAAGTGGAACTATCCAACTGGTGTAT 4683
Db 1 ATGCTTTATGGAGTGGCTCAGATCTTTTGGAGAACTCGACCTGTCCAGCATGGTGTAT 60
QY 4684 TGGATGTTCAACTCTTCCT 4743
Db 61 CGGATGTTCAAAATGTTCCACCGCTCTCACTGGTGGATCCACACTCACTCCCTCTCAC 120
QY 4744 AAGAGAGCTTCCAACTCGTCTCTGGAAGTCTTACCGGACCTCTCTCTCTCTCTCTCT 4803
Db 121 CCGCGCGGCTTCCAGTCTCTCTGGAAGTCTTCTGGAAGTCTTCTGGAAGTCTTCTGGA 180
QY 4804 G 4804
Db 181 G 181

RESULT 6

US-10-027-632-251328/c
; Sequence 251328, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.

; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: Polymorphisms in the Human Genome
; CURRENT APPLICATION NUMBER: US/10/027,632

; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 251328

; LENGTH: 642

; TYPE: DNA

; ORGANISM: Human

US-10-027-632-251328

Query Match 1.9%; Score 96.2; DB 13; Length 642;
Best Local Similarity 83.4%; Pred. No. 3.6e-18;
Matches 121; Conservative 0; Mismatches 23; Indels 1; Gaps 1;

QY 4086 CCGAAGCAGGAGTAAAGACTACGGAGCACTGTTTCAGAGAGACACGAGACCGGGCTAG 4145
Db 610 CAGAGCAGGAGTAAAGAA-ATAAGGAGCACTGTTCCAAAGAGTACAGAAACAGGCTGG 552
QY 4146 CAGTGGAGATGAGAACTGGATGACCCCGCAGCGCCGCGGGAATCCACAGATGGCAGCA 4205
Db 551 CCGTGGAAATGAGAACTGGATGACTCCACAGCGCAAGCCGAGAGTCTACAGATGGTAGCA 492
QY 4206 TGAACAGCTATAGCTCGGAAGGAAA 4230
Db 491 TGAACAGCTACAGCTCAGAAAGGAAA 467

RESULT 7

US-10-027-632-251328/c
; Sequence 251328, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: Polymorphisms in the Human Genome
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 251328
; LENGTH: 642
; TYPE: DNA


```

; LENGTH: 104
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL035633.18
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.81
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1
; OTHER INFORMATION: NT HIT: g114755927, EVALUE 6.00e-52
; OTHER INFORMATION: SWISSPROT HIT: P38552, EVALUE 8.00e+00
; OTHER INFORMATION: EST_HUMAN HIT: AW956936.1, EVALUE 9.00e-52
US-10-029-386-17280

Query Match
Best Local Similarity 1.1%; Score 52.8; DB 13; Length 104;
Matches 72; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 4694 AAATCTTCCCTCCCTCTCCCTCTAGTAGATCCAACTCCGGCACCTCTGACAAAGAGCT 4753
DB 1 AAATGTTCCACCGTCTCTCATCTGTTGGATCCACACTCACTCCCTCACCCTGGGGCT 60

QY 4754 TCCCAATCTCTCTGAAAGTTCTACCGACCTTTTACTCTCG 4797
DB 61 TCCCAATCTCTCTGAAAGTTCTACCGACCTTTTACTCTCG 104

RESULT 15
US-09-908-975-10530
; Sequence 10530, Application US/09908975
; Publication NO. US20030165843A1
; GENERAL INFORMATION:
; APPLICANT: SHOSHAN, Avi
; APPLICANT: WASSERMAN, Alon
; APPLICANT: MINTZ, Eli
; APPLICANT: MINTZ, Liat
; APPLICANT: FAIGLER, Simchon
; TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLICER
; FILE REFERENCE: 36688-0005
; CURRENT APPLICATION NUMBER: US/09/908,975
; CURRENT FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: US 60/287,724
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: US 60/221,607
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 32337
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10530
; LENGTH: 60
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-908-975-10530

Query Match
Best Local Similarity 1.0%; Score 50.4; DB 13; Length 60;
Matches 54; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 3108 CTCATCGAGTAGATTTATAGGAGCAAGCTCATGCTGCGCTAGTGCCTCTCTCTCTC 3167
DB 1 CTCATCGAGTAGATTTATAGGAGCAAGCTCATGCTGCGCTAGTGCCTCTCTCTCTC 60

Search completed: December 12, 2003, 07:10:50
Job time : 2668 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 11, 2003, 15:31:10 ; Search time 8939 Seconds
(without alignments)
13540.250 Million cell updates/sec

Title: US-09-617-099b-2

Perfect score: 4980

Sequence: 1 gcttcctagggtggttcgs.....acattgtgccttagcaagg 4980

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estnu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_esti:*

10: gb_estc:*

11: gb_htc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pln:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_mam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rod:*

26: em_gss_pig:*

27: em_gss_vri:*

28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	937	18.8	1346	11 AK083172	AK083172 Mus muscu
2	891.6	17.9	3687	11 AK032619	AK032619 Mus muscu
3	758.2	15.2	1200	11 BC018057	BC018057 Homo sapi
4	682.4	13.7	686	10 BB428808	BB428808 BB428808

5	678	13.6	739	14	CB519497
6	675.6	13.4	1073	13	BUS02896
7	669.8	13.4	698	14	CB527998
8	588.8	11.8	665	10	BG296106
9	580.4	11.7	594	13	BUS610425
10	546.6	11.0	697	13	BUT41736
11	537.6	10.8	944	13	BUT40396
12	493.8	9.9	603	2	HS0067824
13	487.8	9.8	637	14	CD237320
14	459.8	9.2	567	13	BUT28456
15	446.4	9.0	590	9	AA476826
16	445.2	8.9	747	12	BT732490
17	435	8.7	754	10	BG207226
18	419.4	8.4	953	13	BUS04624
19	416.8	8.4	523	9	AA602733
20	409.2	8.2	502	10	BES01603
21	405.4	8.1	1071	13	BX345439
22	403.6	8.1	588	10	BG403239
23	396	8.0	939	13	BX331506
24	395.6	7.9	474	9	AA772858
25	393.8	7.9	934	13	BUS30896
26	391.4	7.9	571	10	BH750318
27	386.4	7.8	1838	11	BC016274
28	386.4	7.8	3080	11	BC037619
29	384	7.7	915	13	BUA13561
30	381.6	7.7	421	13	BY617064
31	380.2	7.6	478	12	BM691959
32	376.8	7.6	540	13	BUS38952
33	376.2	7.6	582	12	B1397396
34	368.4	7.4	537	13	BX104117
35	368.2	7.4	577	13	BUT11560
36	367.8	7.4	1684	11	BC024828
37	366.6	7.4	960	10	BH791480
38	364	7.3	987	13	BX389364
39	361.6	7.3	1718	11	AK043152
40	361	7.2	1004	14	CD301064
41	354.4	7.1	718	10	BG706509
42	352.2	7.1	600	14	W22505
43	352.2	7.1	937	9	AL519002
44	347.6	7.0	740	10	BG404784
45	347	7.0	359	10	BH864173

ALIGNMENTS

RESULT 1	AK083172	1346 bp	mRNA	linear	HTC 05-DEC-2002
LOCUS	Mus musculus adult male hippocampus cDNA, RIKEN full-length enriched library, clone:C630023J21 product:Rab3 interacting protein 1, full insert sequence.				
DEFINITION	AK083172.1 GI:26350308				
ACCESSION	AK083172				
VERSION	AK083172.1				
KEYWORDS	HTC; CAP trapper.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	Carninci, P. and Hayashizaki, Y.				
TITLE	High-efficiency full-length cDNA cloning				
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)				
MEDLINE	99279253				
PUBMED	10349636				
REFERENCE	2				
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.				
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes				
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)				
MEDLINE	20499374				
PUBMED	11042159				

REFERENCE
AUTHORS

3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsunai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kasaiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, S., Matabiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE
AUTHORS

4 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,
Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamana, I.,
Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,
Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,
Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,
Kuehl, P., Lewis, S., Matsuo, Y., Nikaide, I., Pesole, G.,
Quackenbush, J., Schriml, L.M., Staubli, F., Suzuki, R., Tomita, M.,
Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,
Baldarelli, R., Barah, G., Blake, J., Boffelli, D., Bojunga, N.,
Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C.,
Claretier, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,
Hofmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P.,
Marchionni, L., Mashima, J., Mazzarelli, J., Mombarts, P., Nordone, P.,
Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
Sato, K., Schonbach, C., Seva, T., Shibata, Y., Storch, K.P., Suzuki, H.,
Toyooka, K., Wang, K.H., Weitz, C., Whittaker, C., Wilming, L.,
Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, J.H., Kohtsuki, S.
and Hayashizaki, Y.
Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE
AUTHORS

5 The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE
AUTHORS

6 (bases 1 to 1346)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (15-APR-2002) Yoshitake Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/
Location/Qualifiers

COMMENT

1009

FEATURES

1009

source

1. 1346
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM,DB:C630023J21"
/db_xref="taxon:10090"
/clone="C630023J21"
/sex="male"
/tissue="hippocampus"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
290..>1345
/note="unmated protein product; Rab3 interacting protein 1
match=993)
putative"
/codon_start=1
/protein_id="BAC38794.1"
/db_xref="GI:26350309"
/translation="MPTFLRQVNCNVLSHFGVSPFPAWLQKLGQTLNARKES
PSVDRQNRREYQSEKEDYQVPSQDTPRSPDYADRESQREPOFYEPKGLNVR
DSNRGRHRSKETYIVDEDVEDRDEYERQREKEYQARYSDENLARYFVKPQYERQ
MRLLHAYSRARHERHSDVLANKLSDLSILLMDRPSQRSVSRRAEMNQSY
SMETRAQGSSTPQTSNHSPTPRSPFLDRPDRADSLRQHLDPSSAVRK
TKREKMETLRLNDSLSDDQSESVRPPPPRPHSKKGGKQVLSSESEELASTPEVT
SCDDVLESESVSEKDS"
BASE COUNT 443 a 275 c 334 g 294 t
ORIGIN

CDS

Query Match 18.8%; Score 937; DB 11; Length 1346;
Best Local Similarity 100.0%; Pred. No. 3.5e-186;
Matches 937; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 848 AGAAAAAGAGTCCATCAGTCTCCAGGAGTCAAAATCGAGATACGAGCAAGTCAGAA 907
DB 410 AGAAAAAGAGTCCATCAGTCTCCAGGAGTCAAAATCGAGATACGAGCAAGTCAGAA 469
QY 908 AGAGAGGACTACTCAGATGATGTTCTTCAGATGTTGATCAATGCTCCATCTCTTCGAT 967
DB 470 AGAGAGGACTACTCAGATGATGTTCTTCAGATGTTGATCAATGCTCCATCTCTTCGAT 529
QY 968 TATCTGATAGACGATCTCAGCGTGAAGCTCAATTTATGAGAACCTGGTCAATTAAAT 1027
DB 530 TATCTGATAGACGATCTCAGCGTGAAGCTCAATTTATGAGAACCTGGTCAATTAAAT 589
QY 1028 TACAGGATTTTAAACAGAGAGGAGCCATAGACATTTCCAAAGAGTATATGTCATGATGA 1087
DB 590 TACAGGATTTTAAACAGAGAGGAGCCATAGACATTTCCAAAGAGTATATGTCATGATGA 649
QY 1088 GATGTCAGAGCAGAGATGATATATGAAGCAAAAGAGAGGAGGAGATATACCGAGCGC 1147
DB 650 GATGTCAGAGCAGAGATGATATATGAAGCAAAAGAGAGGAGGAGATATACCGAGCGC 709
QY 1148 TACAGAGTCTCAATCTGCGCGGTATCCCTTAAGCCCAACCTTACCAAGAACAA 1207
DB 710 TACAGAGTCTCAATCTGCGCGGTATCCCTTAAGCCCAACCTTACCAAGAACAA 769
QY 1208 ATGCGCATCCACGCTGAGGTGTCAGGCGCAGCAGATCAGAGAGAGGAGGAGTGTCT 1267
DB 770 ATGCGCATCCACGCTGAGGTGTCAGGCGCAGCAGATCAGAGAGAGGAGGAGTGTCT 829
QY 1268 TTGCGCAACCGCTGAACTAGAGATTTCCAGGATTTCTCTGCTTAAGAGTGTGATAGCAATCA 1327
DB 830 TTGCGCAACCGCTGAACTAGAGATTTCCAGGATTTCTCTGCTTAAGAGTGTGATAGCAATCA 889
QY 1328 AGCGAAGATCTGTATCTGAACTGAGCTGAGTGTGATGAGAAACCAACCAACGATTTCAATG 1387
DB 890 AGCGAAGATCTGTATCTGAACTGAGCTGAGTGTGATGAGAAACCAACCAACGATTTCAATG 949
QY 1388 GAAAGAACTCGAGAGGCTCAGGAGGAGCAAGTCTTATCCCAAGAGGAGCTTCAATCATAGT 1447
DB 950 GAAAGAACTCGAGAGGCTCAGGAGGAGCAAGTCTTATCCCAAGAGGAGCTTCAATCATAGT 1009

QY 1448 CTCTCCACCCCTCGGGGAGCCCTATACCGCTTGATAGACCCAGCATGAGGCGCGCTGAC 1507
 Db 1010 CTCTCCACCCCTCGGGGAGCCCTATACCGCTTGATAGACCCAGCATGAGGCGCGCTGAC 1069
 QY 1508 TCCCTACGGAAACAGACCACTAGATCCAGCTCTGCTGTGAGGAAACACAGACGAGAA 1567
 Db 1070 TCCCTACGGAAACAGACCACTAGATCCAGCTCTGCTGTGAGGAAACACAGACGAGAA 1129
 QY 1568 AAATGGAACCACTGTTAAGGAATGATCTTTGAGTTCAGACAGTCCGAGTCCGAGGAG 1627
 Db 1130 AAATGGAACCACTGTTAAGGAATGATCTTTGAGTTCAGACAGTCCGAGTCCGAGGAG 1189
 QY 1628 CGGCCCCACCAAGGCTCTAATATCCAGAAAGGAGTAAATGCGCAGCTTTCACTG 1687
 Db 1190 CGGCCCCACCAAGGCTCTAATATCCAGAAAGGAGTAAATGCGCAGCTTTCACTG 1249
 QY 1688 AGCAGCTCGGAGGAGGCTGGCATCCACCTGATATACAGCTGTGATGATGAGGAG 1747
 Db 1250 AGCAGCTCGGAGGAGGCTGGCATCCACCTGATATACAGCTGTGATGATGAGGAG 1309
 QY 1748 CTGGAAGCGAGAGTGTGAGTGGAAAGGGGACAGTC 1784
 Db 1310 CTGGAAGCGAGAGTGTGAGTGGAAAGGGGACAGTC 1346

RESULT 2
 AK032619
 LOCUS
 DEFINITION
 Mus musculus adult male olfactory brain cDNA, RIKEN full-length enriched library, clone:6430709C16 product:Rab3 interacting protein 1, full insert sequence.

ACCESSION
 AK032619
 VERSION
 AK032619.1 GI:26328428
 KEYWORDS
 HTC; CAP trapper.
 SOURCE
 Mus musculus (house mouse)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
 1 Carninci, P. and Hayashizaki, Y.
 Title
 High-efficiency full-length cDNA cloning
 Journal
 Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE
 99279253
 PUBMED
 10349636

REFERENCE
 2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Title
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
 Journal
 Genome Res. 10 (10), 1617-1630 (2000)
 MEDLINE
 20499374
 PUBMED
 11042159

REFERENCE
 3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, G., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 Title
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
 Journal
 Genome Res. 10 (11), 1757-1771 (2000)
 MEDLINE
 20530913
 PUBMED
 11076851

REFERENCE
 4 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saico, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, M., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G.,

Quackenbush, J., Schriml, L.M., Staubli, P., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M.P., Brownstein, M.J., Buit, C., Clatc, C., Fuchter, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Rume, D.A., Kamilya, M., Lee, N.H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seva, T., Shibata, Y., Storch, K.P., Suzuki, H., Toyooka, K., Wang, K.H., Weitz, C., Whittaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kotsuki, S. and Hayashizaki, Y.
 Functional annotation of a full-length mouse cDNA collection
 Nature 409 (6821), 685-690 (2001)
 MEDLINE
 21085660
 PUBMED
 11217851

5 The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
 Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
 MEDLINE
 121217851

6 (bases 1 to 3687)
 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, M., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Hori, P., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koyama, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohmoto, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, P., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
 Direct Submission
 Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration and Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
 Please visit our web site for further details.
 URL: http://genome.gsc.riken.go.jp/
 URL: http://fantom.gsc.riken.go.jp/
 Location/Qualifiers
 1..3687
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="PANTOM:DB:6430709C16"
 /db_xref="taxon:10090"
 /clone="6430709C16"
 /sex="male"
 /tissue type="olfactory brain"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev stage="adult"
 221..1078
 /note="unnamed protein product; Rab3 interacting protein 1 (MGD) [GI:2152972, GB|NM_053271, evidence: ELASTN, 100%, match=893]
 putative"
 /codon_start=1
 /protein_id="BAC27953.1"
 /db_xref="GI:26328429"
 /translation="MGRQGLGCTGAAGRSQSRSLGASFEALGYPFWNSLE
 DEGRGCKLRLSTVQRSTETCTGLAVENWNTQASRSTDSGNSYSSEGNLIFVGR


```

Db      637  GAAAGAACTCGAGAGGCTCAGGACAAATTTCTATCCAGAGGACCTCAATATAGT 696
QY      1448  CTCCCAACCCCTCGGCGAGCCCTATACCGCTTGATAGACCAAGATGAGGCGCGCTGAC 1507
Db      697  CCT-CCACCCCTCGGCGAGCCCTATACCGCTTGATAGACCAAGATGAGGCGCGCTGAC 755
QY      1508  TCCTACGGAACAGACCACTTAGATCCAGCTCTGCTGTGAGG 1552
Db      756  TCCCTACGGAACAGCA-CACCTAGATCCAGCTCTGCTGTGAGG 799

RESULT 6
LOCUS    BU502896
DEFINITION AGENCOURT_8929724 NIH_MGC_94 Mus musculus cDNA clone IMAGE:6489689
VERSION   BU502896
KEYWORDS  EST.
SOURCE    Mus musculus (house mouse)
ORGANISM  Mus musculus
REFERENCE 1 (bases 1 to 1073)
AUTHORS   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
JOURNAL   NIH-MGC http://mgc.nci.nih.gov/
COMMENT   National Institutes of Health, Mammalian Gene Collection (MGC)
          Unpublished
          Contact: Robert Strausberg, Ph.D.
          Email: cgsbbs-remail.nih.gov
          Tissue Procurement: The Cepko Laboratory
          cDNA Library Preparation: Life Technologies, Inc.
          DNA Sequencing By: Agencourt Bioscience Corporation
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LLAM14038 row: e column: 18
          High quality sequence stop: 571.
          Location/Qualifiers
            1..1073
              /organism="Mus musculus"
              /mol_type="mRNA"
              /db_xref="taxon:10090"
              /clone="IMAGE:6489689"
              /tissue_type="retina"
              /lab_host="DH10B (phage-resistant)"
              /clone_lib="NIH MGC 94"
              /note="Organ: eye; Vector: pCMV-SPORT6; Site 1: NotI;
              Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
              Average insert size 3.3 kb. Library enriched for
              full-length clones and constructed by Life Technologies.
              Note: this is a NIH MGC Library."
BASE COUNT 340 a 238 c 277 g 217 t 1 others
ORIGIN

Query Match 13.6%; Score 675.6; DB 13; Length 1073;
Best Local Similarity 95.2%; Pred. NO. 3e-131;
Matches 707; Conservative 0; Mismatches 35; Indels 1; Gaps 1;

QY      1091  GTGGAGACAGATGATATGAAGACAAAGGAGGAGGAGGATACAGGACGCTAC 1150
Db      1  GTGGAGACAGATGATATGAAGACAAAGGAGGAGGAGGATACAGGACGCTAC 60
QY      1151  AGAAGTGATCCAAATCTGGCCGGTATCCCTAAAGCCACACCTTACGAGAACAAATG 1210
Db      61  AGAAGTGATCCAAATCTGGCCGGTATCCCTAAAGCCACACCTTACGAGAACAAATG 120
QY      1211  CGCATCCACGCTGAGGTTCCAGGCGCAGCATGAGAGAGGCGCAGTGATGTTCTTTG 1270
Db      121  CGCATCCACGCTGAGGTTCCAGGCGCAGCATGAGAGAGGCGCAGTGATGTTCTTTG 180
QY      1271  GCMAACGCTGAACCTAGAGATTCAGGATTTCTCTGCTAAGATGGATAGACCATCAAG 1330

```

```

Db      181  GCAAAAGCTGAATCAGAGATTCAGGATTTCTCTGCTAAGGATGATAGACCATCAGG 240
QY      1331  CAAAGATCTGATCTGAACGTAGAGCTGCAATGCGAAACCAACGATCTATTTCATAGGAA 1390
Db      241  CAAAGATCTGATCTGAACGTAGAGCTGCAATGCGAAACCAACGATCTATTTCATAGGAA 300
QY      1391  AGAATCTGAGAGGCTCAGGAGCAAAAGTTCTTATCCACAAAGCACTCAATATCATAGTCT 1450
Db      301  AGAATCTGAGAGGCTCAGGAGCAAAAGTTCTTATCCACAAAGCACTCAATATCATAGTCT 360
QY      1451  CCCACCCCTCGGCGAGCCCTATACCGCTTGATAGACCAAGATGAGGCGCGCTGACTCC 1510
Db      361  CCCACCCCTCGGCGAGCCCTATACCGCTTGATAGACCAAGATGAGGCGCGCTGACTCC 420
QY      1511  CTACCGAAACAGACCACTTAGATCCAGCTCTGCTGTGAGGAGGAGGAGGAGGAGGAGG 1570
Db      421  CTACCGAAACAGACCACTTAGATCCAGCTCTGCTGTGAGGAGGAGGAGGAGGAGGAGG 480
QY      1571  ATGGAACCATGTTAAGGAATGATTTCTTGGATTCAGACCTCCAGTCCAGTGTGAGGCGG 1630
Db      481  ATGGAACCATGTTAAGGAATGATTTCTTGGATTCAGACCTCCAGTCCAGTGTGAGGCGG 540
QY      1631  CCCACCAAGAGCTCTATAAATCCAAAGAGAGGATTAATGCGCCAGGTTTCACTGAGC 1690
Db      541  CCCACCAAGAGCTCTATAAATCCAAAGAGAGGATTAATGCGCCAGGTTTCACTGAGC 600
QY      1691  AGCTCGAGGAGG-GGAGCTGGCATCCACACCTGATACAGCTGTGATGATGATGATGAGCT 1749
Db      601  AGCTCGAGGAGGAGGAGCTGGCATCCACACCTGATACAGCTGTGATGATGATGATGAGCT 660
QY      1750  GGAAGCGAGAGTGTGAGTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1809
Db      661  GGAAGCGAGAGTGTGAGTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 720
QY      1810  GCAGGAGTTTGTGCGATTCTA 1832
Db      721  ATCGCCTAAATTTGGCGTATTTTA 743

RESULT 7
LOCUS    CB527998
DEFINITION UI-M-FY0-cfk-j-20-0-UI.r1 NIH BMAP_FY0 Mus musculus cDNA clone
IMAGE: 6850773 5', mRNA sequence.
ACCESSION CB527998
VERSION    CB527998.1 GI:29361471
KEYWORDS  EST.
SOURCE    Mus musculus (house mouse)
ORGANISM  Mus musculus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
          1 (bases 1 to 698)
          NIH-MGC http://mgc.nci.nih.gov/
          National Institutes of Health, Mammalian Gene Collection (MGC)
          Unpublished
          Contact: Robert Strausberg, Ph.D.
          Email: cgsbbs-remail.nih.gov
          Tissue Procurement: Dr. Jim Lin, University of Iowa
          cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
          DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
          DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
          Clone Distribution: Distribution information can be found at
          http://genome.uiowa.edu/distribution/mousefl.html
          This clone was contributed by the Brain Molecular Anatomy Project
          (BMAP)
          Seq primer: pYX-5.
          Location/Qualifiers
            1..698
              /organism="Mus musculus"
              /mol_type="mRNA"
              /strain="C57BL/6"
              /db_xref="taxon:10090"

```


Db 300 AGATGAACCTGAACTATCAACATGGTGATTGGATGTTCAAACTCTTCCCTCTCCCTC 359
 QY 4714 CCTAGTAGATCCCACTCGGCACCTCTGACAGAGAGCTTCCCAATGCTCTGGAAG 4773
 Db 360 CCTAGTAGATCCCACTCGGCACCTCTGACAGAGAGCTTCCCAATGCTCTGGAAG 419
 QY 4774 TTCTACCGGACCTCTTACTCTCTGTTCAATAGCACTATATAAACTGTTGTCAACACCA 4833
 Db 420 TTCTACCGGACCTCTTACTCTCTGTTCAATAGCACTATATAAACTGTTGTCAACACCA 479
 QY 4834 GCGATA-CAAAAACCAAGAGAAACGCAAGGTGGAGCCCTGCTGTAACACTGCTGCTT 4892
 Db 480 GCGATACCAAAAACCAAGAGAAACGCAAGGTGGAGCCCTGCTGTAACACTGCTGCTT 539
 QY 4893 GATGTTGTTCTACAGAGCCCACTCTAGGATACCAAGCACTCTGTTCTTCAAGGA 4952
 Db 540 GATGTTGTTCTACAGAGCCCACTCTAGGATACCAAGCACTCTGTTCTTCAAGGA 599
 QY 4953 AGTCGTACACATGTCCTCTAGCAAGG 4980
 Db 600 AGTCGTACACATGTCCTCTAGCAAGG 627

RESULT 9
 BU610425
 LOCUS
 DEFINITION UI-M-DJ2-bwJ-p-02-0-UI.r1 NIH BMAP DJ2 Mus musculus cDNA clone.
 ACCESSION BU610425
 VERSION BU610425
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 594)
 AUTHORS Ronaldo, M.F., Lennon, G. and Soares, M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 PUBMED 8889548
 COMMENT Contact: Chin, H
 National Institute of Mental Health
 6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
 20892-9643, USA
 Tel: 301 443 1706
 Fax: 301 443 9890
 Email: m85@mail.nih.gov
 Tissue Procurement: Dr. Robin Davisson
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).
 Seq primer: M13 REVERSE.

FEATURES

Location/Qualifiers
 1..594
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="UI-M-DJ2-bwJ-p-02-0-UI"
 /tissue_type="subfornical organ and postrema"
 /dev_stage="Adult"
 /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
 /clone_lib="NIH BMAP DJ2"
 /note="Organ: Brain; Vector: p7T3-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; UI-M-DJ2 is a subtracted cDNA library constructed according to Ronaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double

stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into p7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is: OCTACATGAT, subfornical organ and area postrema."

BASE COUNT 162 a 151 c 144 g 136 t 1 others
 ORIGIN

Query Match 11.7%; Score 580.4; DB 13; Length 594;
 Match Local Similarity 99.7%; Pred. No. 2.8e-111;
 Matches 592; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
 QY 4383 TGGAGGTAGCAATATCAATCCGGGCGCGGCGCTTGT-GGTAAACACCGAGTTCCAGACACTG 4441
 Db 1 TGGAGGTAGCAATATCAATCCGGGCGCGGCGCTTGTGTTAAACACCGAGTTCCAGACACTG 60
 QY 4442 CAGACACCTATGTCTCAAGGTGTATCTGTAGACACGAGTCTGCTATAGCCCAAAAGAAA 4501
 Db 61 CAGACACCTATGTCTCAAGGTGTATCTGTAGACACGAGTCTGCTATAGCCCAAAAGAAA 120
 QY 4502 ACCAGGTGGCAGAAAGACCCCTGGAGCCCTGTATACAGACCTCTTGTCTTCGAGGAG 4561
 Db 121 ACCAGGTGGCAGAAAGACCCCTGGAGCCCTGTATACAGACCTCTTGTCTTCGAGGAG 180
 QY 4562 AGCCCCCAGGGGAGGAGTGTACAGATCATCTCTGCGGAGATATGCTGATGATCAG 4621
 Db 181 AGCCCCCAGGGGAGGAGTGTACAGATCATCTCTGCGGAGATATGCTGATGATCAG 240
 QY 4622 AAATCTTTATGGGAGTGGCCAGATACTCTTTAGATGAATCGAATCTTCAACATGTTG 4681
 Db 241 AAATCTTTATGGGAGTGGCCAGATACTCTTTAGATGAATCGAATCTTCAACATGTTG 300
 QY 4682 ATTGGATGTTCAAACTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4741
 Db 301 ATTGGATGTTCAAACTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 360
 QY 4742 ACAGAGAGCTTCCCAATCTCTCTGGAAGTCTTACCGACCTTCTTCTCTCTCTCTCA 4801
 Db 361 ACAGAGAGCTTCCCAATCTCTCTGGAAGTCTTCTCTCTCTCTCTCTCTCTCTCTCTCA 420
 QY 4802 TAGCAACTATAAACTGTTGTCAACAAACACGAGCATACATAAAACACAGAGAAACGAC 4861
 Db 421 TAGCAACTATAAACTGTTGTCAACAAACACGAGCATACATAAAACACAGAGAAACGAC 480
 QY 4862 AGGTGGAAGCCCTGTTAACTGTCATCTGATGTTGTGTCTACAGAGCCCACTCTAG 4921
 Db 481 AGGTGGAAGCCCTGTTAACTGTCATCTGATGTTGTGTCTACAGAGCCCACTCTAG 540
 QY 4922 GGATACCAAGCAGTCTGTTCTCAGAGAGTCTTACACATTTGTCCTTACG 4975
 Db 541 GGATACCAAGCAGTCTGTTCTCAGAGAGTCTTACACATTTGTCCTTACG 594

RESULT 10
 BU741736/c
 LOCUS BU741736
 DEFINITION UI-E-B01-aiz-h-20-0-UI.s1 UI-E-B01 Homo sapiens cDNA clone
 ACCESSION BU741736
 VERSION BU741736
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 697)
 AUTHORS Ronaldo, M.F., Lennon, G. and Soares, M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477

8889548
PUBMED
COMMENT

Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road , 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
Seq primer: ML3 FORWARD
POLYA=yes.

FEATURES
source

Location/Qualifiers
1. .697
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-E-EOI-aiz-h-20-0-UI"
/tissue_type="fetal eye"
/dev_stage="fetal"
/lab_host="DH108 (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-E-EOI"
/note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; UI-E-EOI is a normalized cDNA library containing the following tissue(s): fetal eye. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CCGGATACCC. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI).
TAG LIB=UI-E-EOI
TAG TISSUE=human fetal eye
TAG_SEQ=CCGGATACCC"
BASE COUNT 164 a 161 c 159 g 212 t 1 others
ORIGIN

Query Match 11.0%; Score 546.6; DB 13; Length 697;
Best Local Similarity 89.4%; Pred. No. 3.6e-104; Mismatches 69; Indels 5; Gaps 3;
Matches 623; Conservative 0;

QY 4165 GATGACCCGCCAGGCGGAGATCCACAGATGGCAGCATGATAGCTC-GG 4223
Db GATGACTCGACAGGCGGAGATCTACAGATGGTACGATGACAGCTCAGCTCAGG 638
QY 4224 AAGGAATCTGATCTTCCCTGGGGTCCGCTGGCTCTGACAGCCAGTTCAGTATCC 4283
Db AAGGAATCTGATCTTCCCTGGGTGTTGGCTTGGCTCTGATAGCCAGTTCAGTATCC 578
QY 4284 TGGATGGCTGGGCTCTGCTAGCTAGTGGGAGCCGACCTGGCTACTCTGCAATGG 4343
Db TGGATGGCTTGGCTCTGCTAGCTAGTGGGAGCCGACCTGGCTAGCAACCTGGATGG 518
QY 4344 GTGACATTCAGTGGGATGATGATGATAA-AAGGACAGCTGGAGGTAGAAATCATCCGG 4402
Db GTGACATTCAGTGGGATGATGATGATAA-AAGGAGGACAGCTGGAGGTAGAAATCATCCGG 458
QY 4403 GGGCGGGCTTGTGGTAAACCAAGTCCAGACACTGCCAGCACCGTATGTCAGGTG 4462
Db GGGCGGGCTTGTGGTAAACCAAGTCCAGACACTGCCAGCACCGTATGTCAGGTG 398
QY 4463 TATCTGTTAGACACGGAGTCTGCATAGCCAAAGAAAAACCAAGGTGGGAGAAAGACC 4522

Db 397 TATCTATTAGATAACGGAGTCTGCATAGCCAAAAAGAAAAACAAGTGCAGAAAAACG 338
QY 4523 CTGGAGCCCTCTGACACAGCTCTTGCTCTCGAGGAGGAGCCGCCACGAGGAGGTGTTA 4582
Db 337 CTGGAAACCTTTTACACAGCTATTATCTTTTGGAGAGAGTCCCAAGGAAGGTTTAA 278
QY 4583 CAGATCATTTGCTGGGAGATTTATGCTGTATGATGATCAAAATCTTTTATGGAGTGGCC 4642
Db 277 CAGATCATCTGCTGGGAGATTTATGCTGTATGATGATCAAAATCTTTTATGGAGTGGCC 218
QY 4643 CAGATCTCTAGATGAATCGAATCTATCCAACTGATGATGATGATGATGATGATGATGAT 4702
Db 217 CAGATCTCTTATGATGAATCTAGAGCTATCCATATGATGATGATGATGATGATGATGAT 158
QY 4703 CCT 4762
Db 157 CCACT 98
QY 4763 TCTCTGAAAGTTTACCGGACCTTTTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4819
Db 97 TCTCTGAAAGTTTCACTGACCTTTTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 38
QY 4820 TGTCAACAACACCGGATGATCAAAACCAAGAAAA 4856
Db 37 TGTCAACAACACCGGATGATCAAAACCAAGAAAA 1

RESULT 11
EUI40296

LOCUS
DEFINITION

ACCESSION
EUI40296

VERSION
EUI40296.1

KEYWORDS
EST.

SOURCE
Gallus gallus (chicken)

ORGANISM
Gallus gallus

REFERENCE
AUTHORS

TITLE
JOURNAL

MEDLINE
PUBMED

COMMENT

603134665F1 CS9CHL24 Gallus gallus cDNA clone ChEST116b17 5', mRNA
sequence.
EUI40296
EUI40296.1 GI:25355131
EST.
Gallus gallus (chicken)
Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 944)
Boardman, P.E., Sanz-Bzquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
A Comprehensive Collection of Chicken cDNAs
Curr. Biol. 12 (22), 1965-1969 (2002)
22335534
12445392

Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology (UMIST)

PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.

FEATURES
source

Location/Qualifiers
1. .944
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="White Leghorn, Hisex"
/db_xref="taxon:9031"
/clone="ChEST116b17"
/dev_stage="16 day embryo"
/lab_host="DH108"
/clone_lib="CS9CHL24"

/note="Organ: brain; Vector: pBluescript II KS(+); Site 1:
EcoRI; Site 2: NotI; Modification of pBluescript II KS(+)
[Stratagene] vector to accommodate cDNA produced with the
T-trimmed protocol (Construction of uni-directionally
cloned cDNA libraries from messenger RNA for improved 3'
end DNA sequencing by Glenn Fu, et al. U.S. Patent # 6,387
, 624). Cut pBluescript II KS(+) with NotI and EcoRI.

QY 4491 CCAAAAGAAAACCAAGTGGGAGAAAGACCCCTGAGAGCCCTGTACACGAGCTCTTGT 4550
 DB 363 CCAAAAGAAAACCAAGTGGGAGAAAGACCCCTGAGAGCCCTGTACACGAGCTCTTGT 304
 QY 4551 CCTTCGAGAGAGCCCGCAGGAGGGGTGTACAGATCAATGTCTGGGAGATTTATGTC 4610
 DB 303 CTTTCAGAGAGTCCACAAGGAAAGTTTTCAGATCATGCTCTGGGAGATTTATGCC 244
 QY 4611 GTATGATCAAAATCTTTATGGAGTGGCCAGAGATCTCTTATGATGAATGGAATAT 4670
 DB 243 GCATGATCAAAATTTTATGGAGTGGCCAGAGATCTTTTATGATGAATGGAATAT 184
 QY 4671 CCAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4730
 DB 183 CCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 124
 QY 4731 CGGCACTCTGACAGAGAGCTTCCCAATCTCTCTGAAAGTTCTACCGAGCTTCTT 4790
 DB 123 TGGCCCTCTGACAGAGAGCTTCCCAATCTCTCTGAAAGTTCTACCGAGCTTCTT 64
 QY 4791 ACTCTGTTCTATGAGCACT---ATAAACTGTTGTCACAAACAGAGATCAAAACC 4847
 DB 63 ACTCTGTTCTATGAGCACTGTTAAATAATTTGTTGTCACAGAACAGGTTACAAAAA 4
 QY 4848 AGA 4850
 DB 3 AAA 1

RESULT 13
 CD237320 697 bp mRNA linear EST 21-MAY-2003
 LOCUS FNPJ04 FNP Homo sapiens cDNA, mRNA sequence.
 DEFINITION CD237320
 VERSION CD237320.1 GI:30980785
 KEYWORDS EST.
 SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 697)
 AUTHORS Ma, Y., Qi, X., Zhang, X., Xiao, H., Zhu, Z., Yan, Q., Chen, Z. and Han, Z.
 TITLE Fetal Pituitary ESTs
 JOURNAL Unpublished
 COMMENT Contact: Zeguang Han
 Chinese National Human Genome Center at Shanghai
 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
 201203, P. R. China
 Tel: 86-21-50801919 (ex.45)
 Fax: 86-21-50801922
 Email: hanzg@chgc.sh.cn.
 Location/Qualifiers
 1. 697
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /tissue_type="pituitary"
 /dev_stage="fetal"
 /clone_lib="FNP"
 /note="Vector: pTriplex2; Site_1: sf11A; Site_2: sf11B"

BASE COUNT 220 a 150 c 201 g 126 t
 ORIGIN
 Query Match 9.8%; Score 487.8; DB 14; Length 697;
 Best Local Similarity 89.5%; Pred. No. 8.3e-92;
 Matches 547; Conservative 0; Mismatches 62; Indels 2; Gaps 2;
 QY 3840 AAATCAGTGTCTTTACATCCAAATGCAAAACAGACAGATGGGCTGTGGGGAGAACT 3899
 DB 89 ATACCAGTGTCTTTACATCCAAATGCAAAACAGACAGATGGGCTGTGGGGAGAACT 148
 QY 3900 TGACAAAGACCAAGCATGATGAGAGATGTGCTCTACAGAGAAATGACGGCAGCC 3959

DB 149 TGACAAAGACCAAGCATGATGAGAGATGT---CTTACTGGAGAGATGATGACGCC 207
 QY 3960 AGTCCGACACCTGCACTGGAGTGGGCGCCCTGGTACAGTGGCAAGCGGCGATTTAGCAT 4019
 DB 208 AGTCTGACACCTGCACTGGAGTGGGCGCCCTGGTACAGTGGCAAGCGGCGATTTAGCAT 267
 QY 4020 GGGCCAAATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4079
 DB 268 GTGCCAAATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 327
 QY 4080 GCCAAACCGAAGGAGGAGTAAAGGCTACGGAGCACTGTTCAGAGAGACGACGAGACCG 4139
 DB 328 GCCAAACCGAAGGAGGAGTAAAGGCTACGGAGCACTGTTCAGAGAGACGACGAGACCG 387
 QY 4140 GGCTAGCAGTGGAGATGAGGAACTGGATGACCCCGAGGCGAGCGGGAATTCACAGAT 4199
 DB 388 GCCTGGCGGTGGAAATGAGGAACTGGATGACCCCGAGGCGAGCGGGAATTCACAGAT 447
 QY 4200 GCAGCATGACAGCTATAGCTTCGGGAGGAAATCTGATCTCCCTGGGTCGCGCTGGGCT 4259
 DB 448 GTAGCATGACAGCTATAGCTTCGGGAGGAAATCTGATCTCCCTGGGTCGCGCTGGGCT 507
 QY 4260 CTGACAGCAGTTCAGTGAATTTCTCTGATGAGGCGCCCTGCTCAGTGTAGTGGAGCC 4319
 DB 508 CTGATAGCCAGTTCAGTGAATTTCTCTGATGAGGCGCCCTGCTCAGTGTAGTGGAGCC 567
 QY 4320 AGACCTGGCTTACTCTGCAATGCGATGAGTGCATTCAGTGGGGAATGATGATAAAGGAC 4379
 DB 568 AGACTCTGGCAACACCTGCAATGCGATGAGTGCATTCAGTGGGGAATGATGATAAAGGAC 627
 QY 4380 AGCTGGAGTGAATCAATCCGGCGCGCGCTGTGTGTTAAACAGGTTCCAGAGCAC 4439
 DB 528 AGCTGGAGTGAATCAATCCGGCGCGCGCTGTGTGTTAAACAGGTTCCAGAGCAC 686
 QY 4440 TGCCAGCACCG 4450
 DB 687 TGCCAGCACCG 697

RESULT 14
 BU728456/c
 LOCUS BU728456
 DEFINITION BU728456 567 bp mRNA linear EST 09-OCT-2002
 UI-B-CL1-aen-b-01-UI.s1 UI-B-CL1 Homo sapiens cDNA clone
 UI-B-CL1-aen-b-01-UI 3', mRNA sequence.
 ACCESSION BU728456
 VERSION BU728456.1 GI:23650352
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 567)
 AUTHORS Ronaldo, M.F., Lennon, G. and Soares, M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene
 discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 PUBMED 889548
 COMMENT Contact: Soares, MB
 Coordinated Laboratory for Computational Genomics
 University of Iowa
 375 Newton Road, 4156 MEBRP, Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: bento-soares@uiowa.edu
 Tissue Procurement: Dr. Gregg Hageman
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research
 Genetics (www.resgen.com).
 Seq primer: M13 FORWARD

POLY(A)-Yes.

FEATURES

source

Location/Qualifiers

1. 567

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="UI-E-CL1-aen-b-01-0-UI"

/tissue_type="human retina"

/dev_stage="adult"

/lab_host="DH10B (Life Technologies) (T1 phage resistant)"

/clone_lib="UI-E-CL1"

/note="Organ: eye; Vector: pTV73-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; UI-E-CL1 is a normalized cDNA library containing the following tissue(s): retina. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directly into pTV73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CCGCG. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI). TAG LIB=UI-E-CL1 TAG TISSUE=human retina TAG_SEQ=CCGCG"

BASE COUNT 133 a 124 c 129 g 181 t

ORIGIN

Query Match 9.2%; Score 459.8; DB 13; Length 567;

Best Local Similarity 89.4%; Pred. No. 6.2e-86;

Matches 507; Conservative 0; Mismatches 57; Indels 3; Gaps 1;

4293 TGGGCGCTGCTCAGCTAGTGGGAGCCGACACCTGGCTACTCTCGAATGGTGACATTC 4352

567 TTGGCCCTGCTCAGCTAGTGGGAGCCGACACCTGGCAACACCTGCGAATGGTGACATTC 508

4353 AGCTGGGAATGATGATATAAAGGAGGACAGCTGGAGGTAGAAATCATCCGGCGCGCGGCC 4412

507 AGGTAGGATGATGAGCAAAAGGAGGACAGCTGGAGGTAGAAATCATCCGGCGCGCGGCC 448

4413 TTGTGTTAAACACAGGTTCCAGACACTGCCAGACCGTATGTCAAGTGTATCTGTAG 4472

447 TTGTTGTTAAACACAGGTTCCAGACACTGCCAGACCGTATGTCAAGTGTATCTGTAG 388

4473 ACAGGGAGTCTGATAGCCAAAGAGAAACCAAGTGGCGGAGGAGACCTGGAGCGCC 4532

387 ATAAAGGAGTCTGATAGCCAAAGAGAAACCAAGTGGCGGAGGAGACCTGGAGCGCC 328

4533 TGTACAGCAGCTCTTGTCTTCGAGGAGAGCGCCCGAGGGAGGGTGTTCAGATCATTTG 4592

327 TTTACAGCAGCTATTTATCTTTGAGAGAGAGTCCACAGGAGAAAGTTTACAGATCATCG 268

4593 TCTGGGAGATTTATGTTCTGATGATCAAAATCTTTATGGAGTGGCCGACATCTCT 4652

267 TCTGGGAGATTTATGCGCATGATCAAAATCTTTTATGGAGTGGCCGACATCTTT 208

4653 TAGATCAATCGGAATCTATCAATGATGATGATGATGATGATGATGATGATGATGATGAT 4712

207 TAGATCAATCGGAATCTATCAATGATGATGATGATGATGATGATGATGATGATGATGAT 148

4713 CCTAGTAGATCAATCCTCGGACCTCTGACAAAGAGGCTTCCCAATGCTCTTGAAA 4772

147 CCTAGTAGATCAATCCTTGGCTCTCTGACAAAGAGGCTTCCCAATCATCTCTGAAA 88

4773 GTTCTACCGGACCTTCTTACTCTCTTCTATAGCACT--ATATAACTGTTGTTCACACA 4829

87 GTTCAACTGGACCTTCTTACTCTCTCTTCTATAGCACTGTAAGAAATTTGTTGTACAGCA 28

4830 ACCAGCGATACAAAACACAGAGAGAAA 4856

DB 27 ACCAGCGTACAAAAAAGAAAAA 1

RESULT 15

LOCUS AA476826

DEFINITION zw95g01.r1 Soares total fetus M02HF8 9w Homo sapiens cDNA clone IMAGE:784752 5' similar to WP:T10A3.1 CE04931 C2 DOMAIN OF PROTEIN KINASE C I, mRNA sequence.

ACCESSION AA476826

VERSION AA476826.1 GI:2205037

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisler, G., Jost, S., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, P., Theising, B., White, Y., Wyllie, T., Waterston, R., and Wilson, R.

AUTHORS WashU-Merck EST Project 1997

TITLE Unpublished

JOURNAL

COMMENT Contact: Wilson RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu This clone is available royalty-free through LML; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Seq primer: -28ml3 rev2 RT from Amersham High quality sequence stop: 495.

FEATURES

Location/Qualifiers

1. 590

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="GDB:5982370"

/db_xref="taxon:9606"

/clone="IMAGE:784752"

/dev_stage="8-9 weeks"

/lab_host="DH10B"

/clone_lib="Soares total fetus M02HF8 9w"

/note="Vector: pTV73-Pac (Pharmacia) with a modified polylinker; Site 1: Site 2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from pooled 8-9 week (total) fetus material with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGAGCGCGCTTAATTTTTTTTTTTT 3'] Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pTV73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 188 a 135 c 129 g 138 t

ORIGIN

Query Match 9.0%; Score 446.4; DB 9; Length 590;

Best Local Similarity 89.4%; Pred. No. 4.1e-83;

Matches 504; Conservative 0; Mismatches 56; Indels 4; Gaps 2;

4296 GCCCTGCTCAGCTAGTGGGAGCGCAGACCTGGCTACTCTCTGCAATGGTGACATTCAGG 4355

5 GCCGTGCTCAGCTAGTGGGAGCGCAGACTCTGGCAACACTGCAATGGTGACATTCAGG 64

4356 TGGGAATGATGATATAAAGGAGCAGCTGGAGGTAGAAATCATTCGGCGCGCGCTTGG 4415

65 TAGGAATGATGACAAAAGGAGCAGCTGGAGGTAGAAATCATTCGGCGCGCGTGCCTTG 123

4416 TGGTAAACACAGGTTCCAGGACACTGGCCGACCGTATGTCAAGTGTATCTGTAGACA 4475

124 TTGTAAACACAGGTTCCAGGACACTGGCCGACCGTATGTAAAGTGTATCTATTAGATA 183

QY 4476 ACCGAGCTCTGATAGCCAAAGAAACCAAGGTCGCGAGAAAGACCTCGGAGCCCTGT 4535
 Db |||||
 QY 184 ACCGAGCTCTGATAGCCAAAGAAACCAAGGTCGCGAGAAAGACCTCGGAGCCCTGT 243
 Db |||||
 QY 4536 ACCGAGCTCTGATAGCCAAAGAAACCAAGGTCGCGAGAAAGACCTCGGAGCCCTGT 4595
 Db |||||
 QY 244 ACCGAGCTCTGATAGCCAAAGAAACCAAGGTCGCGAGAAAGACCTCGGAGCCCTGT 303
 Db |||||
 QY 4596 GGGGAGATTATGGTCGTATGGATCAAAATCCTTTATGGGAGTGCCCGAGATCTTTAG 4655
 Db |||||
 QY 304 GGGGAGATTATGGTCGTATGGATCAAAATCCTTTATGGGAGTGCCCGAGATCTTTAG 363
 Db |||||
 QY 4656 ATGAACTGGAATATCGGAGATCGGAGATCGGAGATCGGAGATCGGAGATCGGAGAT 4715
 Db |||||
 QY 364 ATGAACTGGAATATCGGAGATCGGAGATCGGAGATCGGAGATCGGAGATCGGAGAT 423
 Db |||||
 QY 4716 TAGTAGATCCCACTCGGAGATCGGAGATCGGAGATCGGAGATCGGAGATCGGAGAT 4775
 Db |||||
 QY 424 TAGTAGATCCCACTCGGAGATCGGAGATCGGAGATCGGAGATCGGAGATCGGAGAT 483
 Db |||||
 QY 4776 CTACCGGACCTTTTACT 4832
 Db |||||
 QY 484 CAACTGGGACCTTTTACT 543
 Db |||||
 QY 4833 AGCGATACAAAACCAAGAGAAA 4856
 Db |||||
 QY 544 AGCGTTACAAAAA 567

Search completed: December 11, 2003, 23:10:52
 Job time : 8950 secs